Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyl transferase (GGT) in serum: results from an IFCC multicenter study

Ferruccio Ceriotti1,*, Joseph Henny2, Josep Queraltó3, Shen Ziyu4, Yeşim Özarda5, Baorong Chen6, James C. Boyd7 and Mauro Panteghini8 on behalf of the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL) and Committee on Reference Systems for Enzymes (C-RSE)

1 Diagnostica e Ricerca S. Raffaele, Istituto Scientifico Universitario S. Raffaele, Milano, Italy
2 Laboratoire de Biologie Clinique, Centre de Medicine Preventive, Vandoeuvre-lès-Nancy, France
3 Servei de Bioquimica, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain
4 National Center for Clinical Laboratory, Beijing Hospital Compound, Beijing, P.R. China
5 Uludag University Medical School, Department of Biochemistry and Clinical Biochemistry, Bursa, Turkey
6 Beijing Aerospace General Hospital, Beijing, P.R. China
7 University of Virginia Health System, Department of Pathology, Charlottesville, VA, USA
8 Centro Interdipartimentale per la Riferibilità Metrologica in Medicina di Laboratorio (CIRME), Università degli Studi, Milano, Italy

Abstract

Background: Aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyl transferase (GGT) measurements are important for the assessment of liver damage. The aim of this study was to define the reference intervals (RIs) for these enzymes in adults, paying attention to standardization of the methods used and careful selection of the reference population.

Methods: AST, ALT and GGT were measured with commercial analytical systems standardized to the IFCC-recommended reference measurement systems. Three centers (two in Italy and one in China) measured their own freshly collected samples; one of these centers also measured frozen samples from the Nordic Countries RI Project and from a Turkish center. RIs were generated using non-parametric techniques from the results of 765 individuals (411 females and 354 males, 18–85 years old) selected on the basis of the results of other laboratory tests and a specific questionnaire.

Results: AST results from the four regions (Milan, Beijing, Bursa and Nordic Countries) were statistically different, but these differences were too small to be clinically relevant. Likewise, differences between the upper reference limits for genders was only 1.7 U/L (0.03 μkat/L), allowing a single RI of 11–34 U/L (0.18–0.57 μkat/L) to be defined. Inter-regional differences were not statistically significant for ALT, but partitioning was required due to significant gender differences. RIs for ALT were 8–41 U/L (0.13–0.68 μkat/L) for females and 9–59 U/L (0.15–0.99 μkat/L) for males, respectively. The upper reference limits for GGT from the Nordic Country population were higher than those from the other three regions and results from this group were excluded from final calculations. The GGT RIs were 6–40 U/L (0.11–0.66 μkat/L) for females and 12–68 U/L (0.20–1.13 μkat/L) for males, respectively.

Conclusions: For AST and ALT, the implementation of common RIs appears to be possible, because no differences between regions were observed. However, a common RI for GGT that is applicable worldwide appears unlikely due to differences among populations.

Keywords: alanine aminotransferase; aspartate aminotransferase; γ-glutamyl transferase; reference intervals.

Introduction

The interpretation of medical laboratory data is a comparative decision-making process, and reference values play an essential role in this process. Although the definition of disease-specific decision limits will likely supersede the use of reference values in the future, at present the definition of reference values is still paramount (1). The introduction by the International Federation of Clinical Chemistry (IFCC) and Laboratory Medicine of newly optimized methods (at 37°C) for the measurement of the catalytic activity of several enzymes (2–4) generated the problem of the redefinition of appropriate reference intervals (RIs). The definition of RIs is a complex process and has been well-described in the recently reviewed Clinical and Laboratory Standards Institute
(CLSI–IFCC document C28-A3 (5). It is difficult for a single clinical laboratory to establish RIs properly, particularly for tests whose results are gender and/or age dependent. However, the increasing availability of commercial assays, traceable to the IFCC reference procedures, is eliminating (or at least greatly reducing) the dependency of test results on the method used for measurement and the need for separate RI studies, unless regional population differences are present (6). Currently, there is only limited evidence that ethnic or behavioral differences produce clinically relevant or even statistically significant differences in reference values. Providing that there are no regional population differences, multicenter studies to develop common RIs provide an ideal solution (7, 8). Such studies can be conducted in two different ways: 1) where each participating laboratory performs its own measurements [as described in paragraph 5.2 of the CLSI–IFCC document (5)] or 2) where each center collects the samples from reference individuals, stores them at an appropriate temperature and ships them to a central laboratory that performs all the measurements. The first approach has the disadvantage of requiring thorough verification of the standardization of measurements performed by each participating laboratory, but uses fresh samples, exactly as in clinical practice. The second approach does not require alignment of measurements performed in different laboratories, but uses frozen samples, thereby introducing an additional step in the pre-analytical phase that may affect the results. Multicenter studies, if properly conducted, allows the identification of population differences, the presence of which, at least for some analytes, has been suggested by Ichihara et al. (9, 10), thus requiring caution in the adoption of common RIs (11).

The aim of this work was to define RIs for aspartate aminotransferase (AST) [L-aspartate: 2-oxoglutarate-amino- transferase, EC 2.6.1.1], alanine aminotransferase (ALT) [L-alanine: 2-oxoglutarate aminotransferase, EC 2.6.1.2] and γ-glutamyl transferase (GGT) [(γ-glutamyl)-peptide: amino acid γ-glutamyltransferase, EC 2.3.2.2] that could be applied by all clinical laboratories using an assay system providing results traceable to the corresponding IFCC primary reference measurement procedure. A related goal was to determine, whether ethnic or other population-based differences for these three enzymes were present that would require development of separate RIs.

Materials and methods

Control materials

Frozen native human samples from a single donation, collected according to the protocol in CLSI C37-A (12) and enriched with human recombinant enzymes, were prepared by MCA laboratories, The Netherlands, as described previously (13). Vials containing these materials at three different concentrations, were used for true- ness verification. They were sent on dry ice to three laboratories included in the Joint Committee for Traceability in Laboratory Medicine (JCTLM) list of reference measurement services for enzymes [Centro Interdipartimentale per la Riferibilità Metrologica (CIRME), Deutsche Gesellschaft für Klinische Chemie und Laboratoriums- medizin (DGKL) and Laboraf] for value targeting. At each center the materials were analyzed in duplicate in three different analytical series (six measurements per center) using the IFCC reference measurement procedures (2–4) with home-brew reagents.

To check the internal variability of assays employed in the participating laboratories, the Bio-Rad Liquid Assayed Multiquant material (QC) (Bio-Rad Laboratories Srl, Segrate, Italy) was used.

Structure of the experiments

Three clinical center laboratories were originally involved in the collection and measurements of reference samples (two in Italy and one in China). Additional samples were obtained frozen (–80°C) from the NOBIDA Biobank [samples from the Nordic Countries Reference Intervals Project – NORIP – (14)] and from a Turkish laboratory (located in Bursa). These samples were analyzed in one of the Italian laboratories that were collecting samples. Analyses were performed on three commercial analytical systems [Siemens Advia 2400, Roche Hitachi Modular and Hitachi 7170A (equivalent to Hitachi 917)] employing the manufacturer’s reagents, calibrators, and control materials for each analyzer and strictly following the manufacturer’s instructions. Compliance with the in vitro diagnostic (IVD) directive was indicated through the CE (“Communautés Européennes”) marking of conformity on assays (15). Reagents for analysis of the transaminases included pyridoxal-5-phosphate.

Standardization of participating laboratories

Each participating laboratory received five frozen 0.5 mL aliquots of “trueueness control material” (TCM) (three levels) and 25 vials per level (two levels) of frozen QC material for evaluation of intra-laboratory variability. The control material was shipped on dry ice. To check analytical trueness, the participating laboratories first performed triplicate measurements of the three TCMs in five different analytical runs (15 results for each TCM). If results were not aligned with the expected values assigned by the reference laboratories, the obtained results were used to recalculate the measuring systems and assure optimal traceability to the corresponding IFCC reference procedures. To calculate the correction factors, a regression equation was calculated comparing the target values assigned by the reference laboratories and the overall means obtained by the routine methods (from 21 to 38 replicates, according to the center performing the analysis). It was decided to neglect the intercepts, forcing the regression line through zero to avoid overcorrection related to extrapolation to zero, when analyzing samples with enzyme activities in the range from approximately 40 U/L to approximately 100 U/L.

QC materials were first measured together with the TCM and then used to check the quality of the analytical measurements during the measurement of samples from reference individuals (duplicate measurements at the beginning and end of each analytical run of about 10 reference samples).

Selection of the reference population

The eligible reference individuals signed an informed consent and underwent an anamnestic interview with a physician using an ‘ad hoc’ questionnaire. Clinical exclusion criteria were the following: presence of diabetes mellitus, myopathies, burns and muscle trauma, hypothyroidism, chronic nephropathy, acute and chronic infection, hepatobiliary disease, use of therapeutic drugs with an effect on serum enzyme activities (i.e., warfarin, antiepileptics, diphenyl-hydantoin, aminopyrin, antidepressants, analgesics, antibiotics), pregnancy, body mass index (BMI) >30 kg/m², alcohol assumption >30 g/day, or heavy exercise in the previous days.
Table 1  Laboratory tests and corresponding thresholds used as exclusion criteria.

<table>
<thead>
<tr>
<th>Laboratory test</th>
<th>Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasma fasting glucose</td>
<td>&gt; 7.0 mmol/L (1.26 g/L)</td>
</tr>
<tr>
<td>Serum creatinine</td>
<td>&gt; 18 µmol/L (2 mg/L) above local upper reference limits</td>
</tr>
<tr>
<td>Serum creatine kinase</td>
<td>&gt; 5 µkat/L (300 U/L)</td>
</tr>
<tr>
<td>Serum C-reactive protein</td>
<td>&gt; 20 mg/L</td>
</tr>
<tr>
<td>Serum uric acid</td>
<td>&gt; 475 µmol/L (80 mg/L)</td>
</tr>
<tr>
<td>Serum triglycerides</td>
<td>&gt; 3.39 mmol/L (3.00 g/L)</td>
</tr>
<tr>
<td>Serum total cholesterol</td>
<td>&gt; 7.76 mmol/L (3.00 g/L)</td>
</tr>
<tr>
<td>Serum albumin</td>
<td>&lt; 32 g/L</td>
</tr>
<tr>
<td>Blood erythrocytes</td>
<td>&lt; 4.0 or &gt; 5.9 mil/µL (males)</td>
</tr>
<tr>
<td></td>
<td>&lt; 3.4 or &gt; 5.2 mil/µL (females)</td>
</tr>
<tr>
<td>Blood haemoglobin</td>
<td>&lt; 120 g/L (males)</td>
</tr>
<tr>
<td></td>
<td>&lt; 110 g/L (females)</td>
</tr>
<tr>
<td>Blood white blood cells</td>
<td>&lt; 3000/µL or &gt; 12,500/µL</td>
</tr>
<tr>
<td>Blood platelets</td>
<td>&lt; 1000/L</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>&lt; 36% or &gt; 52% (males)</td>
</tr>
<tr>
<td></td>
<td>&lt; 33% or &gt; 47% (females)</td>
</tr>
<tr>
<td>Mean cellular volume</td>
<td>&lt; 70 fL or &gt; 97 fL</td>
</tr>
<tr>
<td>Hepatitis B (HB) surface</td>
<td>Positive</td>
</tr>
<tr>
<td>Antigens anti-HB core antigen</td>
<td>Positive</td>
</tr>
<tr>
<td>Anti-hepatitis C virus antibodies</td>
<td>Positive</td>
</tr>
</tbody>
</table>

Blood was collected from each individual after an overnight fast and serum was obtained after centrifugation. In addition to the aliquot for the RI study, other blood samples were collected to exclude known causes of AST, ALT, or GGT increases (e.g., subclinical liver or muscle disease) using other laboratory investigations. The tests performed and their cut-off thresholds for subject exclusion are shown in Table 1.

Statistical evaluation

The results from reference individuals were analyzed using RefVal 4.0 software (16). Based on previous data in the literature indicating gender differences (17, 18) for the enzyme catalytic activities evaluated in serum, data for males and females was analyzed separately. A preliminary evaluation for the presence of outliers was performed, but it was decided not to exclude any observation unless clear analytical or biological reasons could be demonstrated.

Results

Value assignment to trueness materials and alignment of laboratory assays

The TCMs were first analysed by reference laboratories. The mean of the means (±SD) of results for TCM levels 1, 2, and 3 obtained in these laboratories were 41.4 U/L (± 1.63 U/L) [0.69 ± 0.027 µkat/L], 70.8 U/L (± 1.69 U/L) [1.18 ± 0.028 µkat/L], and 97.3 U/L (± 1.84 U/L) [1.622 ± 0.031 µkat/L] for AST, 39.7 U/L (± 2.18 U/L) [0.662 ± 0.036 µkat/L], 67.6 U/L (± 2.61 U/L) [1.127 ± 0.044 µkat/L], and 96.9 (± 2.85 U/L) [1.615 ± 0.048 µkat/L] for ALT, 38.3 U/L (± 0.90 U/L) [0.638 ± 0.015 µkat/L], 75.8 U/L (± 1.74 U/L) [1.263 ± 0.029 µkat/L], and 96.6 U/L (± 0.87 U/L) [1.61 ± 0.015 µkat/L] for GGT, respectively. These results were used as target values for standardization purposes of the participating laboratories that measured the same TCMs for verification of the traceability using their commercial assay systems. Figure 1 shows the results obtained compared with the results obtained by the reference laboratories.

As can be seen for ALT, the overlap between the commercial systems and reference procedure values was almost complete; only the high-level TCM in Beijing was outside the limits, but since the discrepancy was seen only for values well above the RI, no data correction was used. For AST, results from Milan 2 did not match the reference procedure intervals, thus that reference values obtained from this laboratory were recalculated by multiplying by a correction factor of 0.933 obtained from the corresponding regression equation. For GGT, the data for all centers were recalculated using the following correction factors: Milan 1 = 0.965, Milan 2 = 1.046, and Beijing = 1.031.

Reference intervals

After application of the exclusion criteria, a total of 810 individuals (142 Asians and 668 Caucasians) were enrolled in the study: 188 (101 females, 87 males) from Milan (two centers), 136 (61 females, 75 males) from Beijing, 356 from NOBIDA Biobank (188 females, 168 males), and 130 from Bursa (79 females, 51 males). However, 45 of these individuals (39 Asians and six Caucasians) were excluded from the calculation because they were younger than 18 years. Characteristics of the four population groups are shown in Table 2, and the distribution of the individuals according to age and ethnicity are shown in Table 3.

For AST and GGT, we observed a weak but statistically significant (p ≤ 0.05) correlation with age and BMI for both genders (r² < 0.09), while ALT was only slightly related to the BMI (r² = 0.031 in females and 0.076 in males, respectively).

Reference individuals

No data was eliminated as statistical outliers. RIs were computed non-parametrically for each center, and the Lahti algorithm (19, 20) was applied to the corresponding upper reference limits to judge, whether partitioning of the data was required. Because the Lahti approach can be applied to a pair of groups only, each regional group was compared with the other three (a total of six comparisons) (Table 4). For none of the three enzymes did the algorithm indicate the need for partitioning of a particular group in all the pair wise comparisons with other groups.

AST

Females  AST activities across the four groups were significantly different (one-way analysis of variance; ANOVA), but the Lahti algorithm indicated no need of partitioning for
Figure 1  Results obtained on the “trueness control material” by the three participating laboratories (means±SD are indicated for each enzyme and at each level). Grey areas between dashed lines represent the mean±SD area obtained with results by the reference laboratories.

Table 2  Characteristics of the reference population groups enrolled in the study.

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th></th>
<th>Males</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of individuals</td>
<td>Age, years^a</td>
<td>BMI, kg/m^2</td>
<td>Number of individuals</td>
</tr>
<tr>
<td>Milan</td>
<td>97</td>
<td>40 (21–72)</td>
<td>21.2 (16.2–28.7)</td>
<td>85</td>
</tr>
<tr>
<td>Beijing</td>
<td>47</td>
<td>47 (21–73)</td>
<td>21.7 (18.0–29.7)</td>
<td>50</td>
</tr>
<tr>
<td>NORIP</td>
<td>188</td>
<td>52 (19–90)</td>
<td>23.1 (17.3–29.8)</td>
<td>168</td>
</tr>
<tr>
<td>Bursa</td>
<td>79</td>
<td>32 (20–54)</td>
<td>22.1 (16.9–30.0)</td>
<td>51</td>
</tr>
</tbody>
</table>

^aMedian (range). BMI, body mass index.

Table 3  Distribution of reference individuals according to age (years) and ethnicity.

<table>
<thead>
<tr>
<th>Age</th>
<th>Males</th>
<th></th>
<th>Females</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Caucasian</td>
<td>Asian</td>
<td>Total</td>
<td>Caucasian</td>
</tr>
<tr>
<td>18–30</td>
<td>70</td>
<td>12</td>
<td>82</td>
<td>96</td>
</tr>
<tr>
<td>31–40</td>
<td>61</td>
<td>11</td>
<td>72</td>
<td>78</td>
</tr>
<tr>
<td>41–60</td>
<td>98</td>
<td>21</td>
<td>118</td>
<td>121</td>
</tr>
<tr>
<td>61–85</td>
<td>72</td>
<td>9</td>
<td>81</td>
<td>68</td>
</tr>
<tr>
<td>Total</td>
<td>301</td>
<td>53</td>
<td>354</td>
<td>363</td>
</tr>
</tbody>
</table>
### Differences between upper reference limits obtained in individuals enrolled in different centers.

<table>
<thead>
<tr>
<th></th>
<th>AST females</th>
<th>AST males</th>
<th>ALT females</th>
<th>ALT males</th>
<th>GGT females</th>
<th>GGT males</th>
</tr>
</thead>
<tbody>
<tr>
<td>Beijing</td>
<td>Np</td>
<td>P</td>
<td>Np</td>
<td>P</td>
<td>Np</td>
<td>P</td>
</tr>
<tr>
<td>Bursa</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>Milan</td>
<td>Np</td>
<td>M</td>
<td>Np</td>
<td>M</td>
<td>M</td>
<td>M</td>
</tr>
<tr>
<td>NORIP</td>
<td>P</td>
<td>M</td>
<td>Np</td>
<td>M</td>
<td>Np</td>
<td>M</td>
</tr>
</tbody>
</table>

The results from each center are compared with the others using the Lahti algorithm to judge the need for partitioning: P, partitioning suggested (more than 4.1% of the population of a single group lies outside the limits described by the distribution of the common group); Np, no need of partitioning (<3.2% of the population of a single group lies outside the limits described by the distribution of the common group); M, marginal (a percentage between 3.2 and 4.1% of the population of a single group lies outside the limits described by the distribution of the common group).

### ALT

#### Females

The Beijing group appeared to be somewhat different from the other regional groups (marginal or partitioning indication according to Lahti, see Table 4). However, because of the limited number of individuals in this group (n=47), we decided to combine all 411 data points. The resulting RIs are shown in Table 5.

#### Males

Two clearly abnormal results were present (one in the Milan group: 113 U/L [0.188 µkat/L] and the second in the NORIP group: 123 U/L [0.205 µkat/L]). Since there were no evident reasons for explaining these results, they were not excluded. Although the upper reference limit of the Beijing group had marginal statistical differences from the other groups, these differences were judged to be clinically irrelevant and all the data was merged (Table 5).

### GGT

#### Females

The 97.5th percentile value for the NORIP group was markedly different from those of other three regions. Even after elimination of an outlier (173 U/L–2.88 µkat/L), the 97.5th percentile value was 68.9 U/L (1.15 µkat/L), whereas the 97.5th percentile values for the other groups ranged from 39.0 U/L (0.65 µkat/L) (Beijing) to 44.1 U/L (0.74 µkat/L) (Milan). Based upon this difference, the NORIP data were excluded from the final calculation, even if the Lahti algorithm indicated a marginal statistical difference (Table 4). The RIs obtained for the remaining 223 individuals are shown in Table 5.

#### Males

As seen for females, the 97.5th percentile value for males in the NORIP group (113.9 U/L [1.90 µkat/L]) was markedly different from those of other three groups which ranged from 64.2 U/L (1.07 µkat/L) (Beijing) to 69.1 U/L (1.15 µkat/L) (Milan). Thus, even for males the RI was calculated by excluding the NORIP data. The resulting reference limits (on 186 individuals) are shown in Table 5.
Table 5  Proposed reference intervals for the three enzymes that were evaluated.

<table>
<thead>
<tr>
<th></th>
<th>Females</th>
<th>Males</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2.5th percentile (90% CI)</td>
<td>97.5th percentile (90% CI)</td>
<td>2.5th percentile (90% CI)</td>
</tr>
<tr>
<td>AST U/L</td>
<td>11.0 (10.0–12.0)</td>
<td>33.4 (30.0–35.5)</td>
<td>13.9 (11.0–14.7)</td>
</tr>
<tr>
<td>µkat/L</td>
<td>0.18 (0.17–0.20)</td>
<td>0.56 (0.50–0.59)</td>
<td>0.23 (0.18–0.25)</td>
</tr>
<tr>
<td>ALT U/L</td>
<td>7.8 (6.0–8.0)</td>
<td>41.0 (35.0–57.0)</td>
<td>9.0 (7.8–10.0)</td>
</tr>
<tr>
<td>µkat/L</td>
<td>0.13 (0.10–0.13)</td>
<td>0.68 (0.58–0.95)</td>
<td>0.15 (0.13–0.17)</td>
</tr>
<tr>
<td>GGT U/L</td>
<td>6.4 (4.8–6.8)</td>
<td>39.7 (32.0–58.0)</td>
<td>11.7 (10.3–12.5)</td>
</tr>
<tr>
<td>µkat/L</td>
<td>0.11 (0.08–0.11)</td>
<td>0.66 (0.53–0.97)</td>
<td>0.20 (0.17–0.21)</td>
</tr>
</tbody>
</table>

*NORIP data excluded. CI, confidence interval.

Discussion

AST, ALT and GGT are among the most commonly requested laboratory tests, and are key for the laboratory evaluation of liver damage. Values above the upper reference limits of these enzymes are often used to trigger further clinical and laboratory investigations. Incorrect setting of these limits can result in costly and invasive evaluations or, conversely, may allow life-threatening liver disease to progress unrecognized. Factors that influence the correctness of these limits can be divided into two categories: 1) differences in analytical methodologies, and 2) differences in the way the RIs were defined. The first factor can be minimized by the adoption of methods producing results traceable to the corresponding reference measurement system (6). The second factor is more relevant, as demonstrated by several authors reporting a wide range of RIs, even in laboratories using similar methods (21–25). Possible reasons to explain this phenomenon are numerous and include differences in the criteria used to select the reference individuals, the number of individuals selected, the statistical approach used to derive the reference limits, or the adoption by laboratories of manufacturers’ suggested RIs, often without any validation. The importance of improving both assay standardization and approaches for defining RIs has been discussed by Wu recently (26).

In our study, we attempted to improve the current situation by obtaining sound reference values, following a well-defined methodology for the selection of the reference individuals and using analytical methods with demonstrated traceability. For effective exclusion of individuals with minor subclinical hepatic damage, such as that related to non-alcoholic fatty liver disease (NAFLD), ultrasound examination should have been performed. However, as screening for reference individuals by ultrasound was not economically feasible, we excluded individuals with conditions known to be correlated with NAFLD: obesity (BMI >30 kg/m²) and hyperlipidemia (total serum cholesterol and triglycerides >3 g/L) (27, 28). Other common causes of liver damage, including excessive alcohol consumption or viral hepatitis B or C, were also excluded. Samples obtained through the NOBIDA Biobank did not have hepatitis B or C testing performed at the origin, but all the samples were tested for hepatitis B surface antigen and hepatitis C antibodies in the center performing the enzyme measurements. All the samples with very high GGT (>80 U/L) were also tested for carbohydrate-deficient transferrin (CDT), with the elimination of any sample with abnormal results (CDT >2%). Thus, our selected population had a low probability of liver abnormalities, even if subclinical cytomegalovirus or Epstein-Barr infection cannot be excluded.

Analytically, the traceability of our results to the reference measurement system was verified through the use of TCMs that had values assigned by the reference procedures. The objective of our work was to define RIs that could be adopted by any laboratory using analytical systems that provide traceable results, but also to verify the possible existence (or not) of differences in the upper reference limits due to ethnicity or life habit. The latter was only partially achieved. In fact, the skewed distribution of the data and the gender-related differences require a larger number of study subjects to check for the possible effects of these factors. GGT in serum was the only enzyme that appeared to be markedly different in the Nordic Countries group vs. the other three regional populations for both males and females. The difference was caused essentially by a small subgroup of individuals with substantially higher GGT values that were not explained by the commonly known factors that increase GGT, such as drug intake (explicitly excluded by the individuals) or chronic alcohol consumption (excluded by CDT measurements). For AST, the differences between groups (regional or gender-related), although sometimes statistically significant, were considered too small to be clinically important. This permitted merging of all the results to derive a single RI. The small difference between the upper reference limits found in the two genders was in agreement with previous literature. Indeed, Schumann and Klauke (29) found a 4 U/L difference (31 vs. 35 U/L) and Leino et al. a 3 U/L difference (33 vs. 36 U/L) (17).

The available literature on ALT reference values is huge, owing to the relevance of this measurement in Transfusion Medicine. The reported upper reference limits for ALT
ranged from 19 U/L and 31 U/L for females and males, respectively, in the report by Prati et al. (30) to 46 U/L and 68 U/L, respectively, in the Nordic Project (31). However, the study of Prati et al. had several different characteristics: data were obtained using an unoptimized assay (no pyridoxal phosphate in the reagent mixture) and the upper reference limit was set at the 95th percentile and not at the 97.5th percentile of the data distribution. Furthermore, all the ALT results > 40 U/L (males) and > 30 U/L (females) were discarded a priori, and the BMI limit for subject inclusion was set at 24.9 kg/m². For all the reported reasons, the limits derived by Prati et al., which were set not to evaluate healthy individuals, but to monitor anti-HCV–positive subjects over time, cannot be compared with those derived in the present study. However, our ALT data are comparable with those obtained by Chan et al. in a Chinese population (upper reference limits of 36 U/L and 53 U/L for females and males, respectively), although they used an assay that did not contain pyridoxal phosphate (32). However our results are substantially higher than those obtained by Brinkmann et al. on blood donors (females, 34.4 U/L; males, 43.9 U/L) (33) and Schumann and Klauke who studied hospitalized patients (females, 34 U/L; males, 45 U/L) (29). Moreover, ALT RIs in our study did not appear to be age-related, contrasting with results shown in some data mining studies where, especially for males, ALT values appear to increase in middle age (34, 35). This difference can likely be explained by our careful selection of individuals, whereas in data mining studies it is more difficult to exclude alcohol and therapeutic drug users.

For AST and GGT, the correlation with age was statistically significant, particularly for women, but it was extremely weak and clinically irrelevant. Also, the effect of BMI, when limited below our pre-defined threshold of 30 kg/m², appeared weak and did not affect the calculation of the upper reference limit. Insufficient data was available for children and adolescents to calculate specific RIs for these age groups. Although a recent publication (36) has found that the RIs for AST, ALT and GGT do not change after the first few years of age, thereafter being very similar to those found in adults, we decided not to include the data of children and adolescents in the calculation of RIs.

GGT results from the Nordic Countries subgroup (upper reference limit of 69 U/L and 114 U/L for females and males, respectively) were similar to the original data published by Strømme et al. for individuals older than 40 years (77 U/L and 114 U/L) (31). This occurred despite the fact that our results were derived from a subset of approximately 350 of the original 1391 participants in the NORIP study, and our selection criteria were more stringent (45 out of 400 samples originally received were excluded primarily for excessive BMI, dyslipidemia, and increased mean erythrocyte volume, two patients also were hepatitis B surface antigen positive, three anti-hepatitis B core antigen positive and three with CDT > 2%). The difference in GGT values between the Nordic Countries group and the other three groups was only marginally significant according to Lahti. Although in each regional population a small subgroup of individuals had higher GGT values, in the Nordic Countries group this subgroup was more numerous and caused a dramatic increase in the upper reference limit (Figure 2). Nevertheless, we were not able to find biological reasons able to explain this behaviour. Perhaps evaluation of the pattern of GGT subfractions in different groups could help in shedding light on this issue (37, 38). The GGT results for females obtained by merging the other three groups (upper limit of [40 U/L (0.67 µkat/L)]) compare well with those shown by Schumann and Klauke (38 U/L) (29) and Leino et al. (42 U/L) (17). For males, the upper reference limit for GGT (68 U/L [1.13 µkat/L]) is higher than that obtained in two studies mentioned previously (55 U/L and 58 U/L, respectively) and also by Steinmetz et al. (49 U/L) (39). However, our GGT data was highly skewed and the 90% confidence interval (CI) around the upper reference limit includes all these previously published limits.

Conclusions

Our study demonstrates that for AST and ALT, common RIs are applicable. The pre-requisite is the use of analytical systems providing results that are traceable to IFCC reference...
measurement systems. If this is the case, the proposed RIs can be adopted, replacing those currently in use. The case of GGT appears to be more complex, and a common RI cannot definitively be suggested, even though three out of four regional reference sample groups (Italy, Turkey and China) showed very similar results. Further studies are needed to understand the reasons of these differences found in apparently healthy individuals, not otherwise affected by any of the known causes that increase GGT.

Acknowledgements

The authors acknowledge Gerhard Schumann (Hannover Medical School, Hannover, Germany), Elena Guerra (Diagnostica e Ricerca San Raffaele, Milano, Italy) and Ilenia Infusino (CIRME, University of Milano, Italy) for enzyme measurements using reference measurement procedures. Alessandra Ratto, Carlo Ferrero (HSR, Milano, Italy) and Cristina Valente (Luigi Sacco University-Hospital, Milano, Italy) are also acknowledged for the invaluable support in selecting reference individuals and measuring corresponding samples. Kiyoshi Ichihara (Yamaguchi University Graduate School of Medicine, Ube, Yamaguchi, Japan) has supervised data elaboration.

Conflict of interest statement

Authors’ conflict of interest disclosure: The authors stated that there are no conflicts of interest regarding the publication of this article. Research funding 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.

Research funding: This study was partially supported by the IFCC. Minor economic support was also given by Abbott Laboratories, Ortho-Clinical Diagnostics Inc., Sysmex Europe GmbH, and The Binding Site Ltd. The Bursa part of the study was supported by The Uludag University Research Fund (Grant # UAP(T)-2009/10).

Employment or leadership: None declared.

Honorarium: None declared.

References