Yesim Ozarda, Kiyoshi Ichihara*, Julian H. Barth and George Klee and on behalf of the Committee on Reference Intervals and Decision Limits (C-RIDL), International Federation for Clinical Chemistry and Laboratory Medicine

Protocol and standard operating procedures for common use in a worldwide multicenter study on reference values

Abstract

The reference intervals (RIs) given in laboratory reports have an important role in aiding clinicians in interpreting test results in reference to values of healthy populations. In this report, we present a proposed protocol and standard operating procedures (SOPs) for common use in conducting multicenter RI studies on a national or international scale. The protocols and consensus on their contents were refined through discussions in recent C-RIDL meetings. The protocol describes in detail (1) the scheme and organization of the study, (2) the target population, inclusion/exclusion criteria, ethnicity, and sample size, (3) health status questionnaire, (4) target analytes, (5) blood collection, (6) sample processing and storage, (7) assays, (8) cross-check testing, (9) ethics, (10) data analyses, and (11) reporting of results. In addition, the protocol proposes the common measurement of a panel of sera when no standard materials exist for harmonization of test results. It also describes the requirements of the central laboratory, including the method of cross-check testing between the central laboratory of each country and local laboratories. This protocol and the SOPs remain largely exploratory and may require a reevaluation from the practical point of view after their implementation in the ongoing worldwide study. The paper is mainly intended to be a basis for discussion in the scientific community.

Keywords: cross-check study; harmonization; multicenter study; panel of sera; reference interval; standardization.

*Corresponding author: Kiyoshi Ichihara, MD, PhD, Department of Clinical Laboratory Sciences, Faculty of Health Sciences, Graduate School of Medicine, Yamaguchi University, Minami-Kogushi 1-1-1, Ube 755-8505, Japan, Phone: +81-836-22-2884, Fax: +81-836-35-5213, E-mail: ichihara@yamaguchi-u.ac.jp
Yesim Ozarda, MD, PhD: Department of Biochemistry and Clinical Biochemistry, Uludag University Medical School of Medicine, Bursa, Turkey
Julian H. Barth, MD, PhD: Blood Sciences, Leeds General Infirmary, Leeds, UK

George Klee, MD, PhD: Laboratory Medicine and Pathology, Mayo Clinic College of Medicine, Rochester, MN, USA

1 Introduction

The interpretation of data in laboratory medicine is a comparative decision-making process, and reference intervals (RIs) given in laboratory reports have an important role in aiding the clinician in interpreting test results in reference to values for healthy populations. Careful determination and/or validation of RIs by the laboratory for use in the patient population it serves is therefore important to ensure their proper utility. About 30 years ago, the International Federation of Clinical Chemistry (IFCC) recommended that each laboratory produce its own reference values and estimate the corresponding RIs according to defined procedures [1–6]. The selection and recruitment of a sufficient number of reference subjects is difficult, time-consuming, and costly. Furthermore, the continued evolution of assay procedures and platforms requires that this process be repeated frequently. The requirement that each clinical laboratory produce its own RIs is practically impossible for most of the clinical laboratories. Thus, although some laboratories have performed local studies for their own use, there have also been multicenter studies performed with recruitment of appreciable numbers of subjects to establish useful RIs by laboratories in the Nordic countries [7], Spain [8], Australia [9], and Asia [10, 11]. Recently, on behalf of the IFCC Committee on Reference Intervals and Decision Limits (C-RIDL), a few multicenter studies have been conducted to obtain well-defined "common" RIs for enzymes such as AST, ALT, and GGT [12]. In addition, the third multicenter Asian study on reference values conducted in 2009 had the following features [13, 14]: (a) a sufficient number of subjects (n=3540) recruited using well-defined criteria, (b) a total of 72 analytes assayed, (c) centralized measurement used to eliminate method-related variations, (d) calibration of RIs for

standardized analytes using certified reference materials (CRMs) or value-assigned sera to ensure traceability to the reference measuring procedure (RMP), (e) cross-check testing between the central laboratory and participating laboratories using a part of sera freshly collected.

Using these same principles, C-RIDL is currently coordinating a multicountry/multicenter RI study with recruitment of a sufficient number of individuals to ensure traceability and harmonization of the test results. C-RIDL expects that the combined results of the studies will reveal a comprehensive picture of reference values and their sources of variations. We may identify regional, ethnic, and lifestyle-related variations in test results after adjusting for assay-platform-dependent factors. However, although we have set clear inclusion criteria, these criteria are essentially pragmatic and there may be subtle differences in interpretation of what constitutes healthiness by different cultures and investigators. We are aware that this may inhibit the derivation of globally applicable common RIs even when we do not observe any region- or ethnicity-related differences.

In any case, the resulting information will contribute to the globalization of medical practice. To achieve the above objectives, it is necessary to prepare well-defined, commonly applicable protocol and standard operating procedures (SOPs), that cover all important aspects, to be used in the multicenter RI studies. The CSLI/IFCC Document C28-A3 contains most of the necessary information related to RI: the selection of reference individuals, preanalytical and analytical conditions, analysis of reference values, transference, RI validation, and medical decision limits, etc. [15]. There are also some proposed guidelines that originated from regional studies [16, 17]. However, these guidelines and the section of Document C28-A3 (Part 6.2) dealing with multicenter studies do not sufficiently cover all the necessary aspects of multicenter studies on reference values.

This proposed protocol was created, and consensus on the contents built through discussions in recent C-RIDL meetings in Anaheim (July 2010), Munich (December 2010), Berlin (May 2011), and Atlanta (July 2011). The protocol covers the following in detail: (1) the scheme and organization of the study, (2) the target population, inclusion/exclusion criteria, ethnicity, and sample size, (3) health status questionnaire, (4) target analytes, (5) blood collection, (6) sample processing, storage, and transportation, (7) assays, (8) cross-check testing, (9) ethics, (10) data analysis, and (11) reporting of results. Three SOPs (SOP 1, SOP 2, and SOP 3) were developed for implementing the study in accordance with the protocol. SOP 1 covers the procedures for recruitment of reference subjects, SOP 2 details the procedures for analysis, and SOP 3 describes data analysis and data reporting.

In the protocol, requirements for conducting the study, phase by phase, are described. Overall, the procedure for standardization of test results is of the utmost importance, and all centers need to comply with it in dealing with standardized analytes. In addition, the protocol proposes the common measurement of a panel of sera in cases where no standardized materials exist for harmonization of test results. It also describes the requirements of the central laboratory, including the method of cross-check testing between the central laboratory of each country and local laboratories [14, 18].

This paper is intended mainly to provide a basis for discussion in the scientific community.

2 Protocol

2.1 The scheme and organization of the multicenter RI study

1. Laboratories of each country will conduct their own multicenter study to derive country-specific RIs and to explore sources of variation for test results relevant to that country. Additional target analytes and questionnaire items can be added according to local needs.

2. Collaborating laboratories in each country should recruit appropriate healthy volunteers, draw blood, and process the specimens observing the common protocol.

3. The centralized assay scheme should be used in each country to eliminate variations due to analytical methods. One or two laboratories may act as central laboratories in each country, receiving specimens from local laboratories. Each central laboratory can use any assay platform for measurement.

4. RIs should be made traceable to the RMPs for standardized analytes through the measurement of standard reference materials (SRMs) or value-assigned sera in collaboration with the Joint Committee for Traceability in Laboratory Medicine (JCTLM).

5. For the nonstandardized analytes, centrally determined RIs should be converted to those of each participating laboratory through cross-checking results with those of the central laboratory using statistical approach such as linear regression [14].

6. In an attempt to compare test results across multiple countries after adjusting for various factors, an approach using a panel of sera (freshly prepared from reference individuals by C-RIDL for the purpose) was

adopted. Using this approach, every participating laboratory would be expected to analyze this panel of sera along with the locally-acquired specimens in order to make the results comparable across the countries [18].

7. A portion of all specimens can be saved for new analytes to be assayed in the future.

2.2 Target population, inclusion/exclusion criteria, ethnicity, and sample size

The following inclusion and exclusion criteria are to be observed in recruiting volunteers.

2.2.1 Inclusion criteria

1. The participants should be feeling subjectively well.
2. The participants should be older than 18 years of age. At least 80% of subjects should be 18–65 years, with equal gender mix and age distributions, except for individuals over 65 years of age.
3. The participants ideally should not be taking any medication. Any subject taking medications or vitamin supplements should have them recorded (name, dose, and frequency) so that secondary exclusion after measurement can be done as required. The following medications are permitted but should be recorded: contraceptive pills or estrogens and thyroxine, if the subject is well replaced (i.e., TSH is lower than the upper reference limit), are permitted, but they should be recorded.

2.2.2 Exclusion criteria

The participant may not enter the study if any of the following applies:
1. Known diabetes on oral therapy or insulin (diet alone is acceptable).
2. History of chronic liver or kidney disease.
3. Blood test results that clearly point to a severe disease.
4. History of being a hospital in-patient or otherwise seriously ill during the previous 4 weeks.
5. Blood donation in the previous 3 months.
6. Known carrier state for HBV, HCV, or HIV.
7. Female participants who are pregnant, breastfeeding, or within 1 year after childbirth.
8. Any other significant disease or disorder that, in the opinion of the investigator, may either put the participants at risk because of participation in the study or may influence the results of the study.
9. Participation in another research study involving an investigational product in the past 12 weeks.

2.2.3 Ethnicity

Information on ethnicity can be collected from the volunteers, if relevant, using the classification shown in Appendix A.

2.2.4 Sample size

The practically attainable target sample size from each country is set at a minimum of 500 (male and female: 250×2) or more, which is greater than twice the minimum number recommended by C28-A3 [120×2 (men and women)], so that country-specific RIs can be obtained in a more reproducible manner. This number is adequate for making between-country comparisons of test results with a power of detecting a difference of two means equivalent to 0.25 times SD comprising the RI (SD$_{RI}$), which corresponds to a bias of 0.25 times between-individual variation, allowing errors of $\alpha<0.05$ and $\beta<0.2$ in the statistical hypothesis testing done separately for each gender. If there is an interest in exploring regional within-country variations, it is recommended to obtain at least 120 (men and women, 60×2) samples from each local area to acquire adequate power to test for a difference of two means equivalent to 0.5×SD$_{RI}$ by the above specification.

2.3 Questionnaire

A sample health status questionnaire is shown in Appendix B. It can be customized to local needs by adding and removing query items. The essential items required for the worldwide comparison are BMI, special diet, records of medicines and/or supplements regularly taken, status of menstruation, habits of smoking, alcohol consumption per week (roughly expressed grams of ethanol), and frequency and strength of physical exercise. The information will be used for analyzing sources of variation in test results and for judging the necessity of a secondary exclusion.

2.4 Target analytes

These are listed in Appendix C.

2.5 Blood collection

The healthy volunteers are strictly requested to avoid excessive physical exertion/exercise for 3 days before sampling. Participants should avoid excessive eating and drinking the night before and should fast overnight for at least 10 h. Sampling should be postponed when subjects are in a state of unusual stress or in the morning immediately after a night shift.

Because the most commonly measured analytes are not influenced by the tube type [19–21], the type, either plain or gel-separator, can be determined by each laboratory. The type of tube and venipuncture equipment used must be recorded.

The amount of blood to be drawn can be determined in each country according to the analytes being tested locally. The time of sampling should be set at 7 to 10 AM. The blood should be drawn after the participant has sat quietly at least for 30 min to avoid variation due to postural influence and physical stress. The waiting time period can be used to fill in and check the health questionnaire.

2.6 Sample processing, storage, and transportation

The blood should be stirred well within the collection tube for balanced clotting. It should be left at room temperature before centrifugation, which should be performed within 1 h. After separation of the serum, the specimen should be promptly divided into aliquots of 1–2 mL using well-sealed freezing containers and be immediately stored at −80°C. All aliquots that have not been taken for local use will be shipped on dry ice to the central laboratory for collective measurement (see the section on SOP 1 below).

The aliquots must be kept at each laboratory to use for cross-check testing to be done at the time of the assays by the central laboratory.

2.7 The centralized measurements

2.7.1 Requirements for the central laboratories

One or two laboratories in each country should be chosen as central laboratories, which will provide collective measurements. Requirements for a central laboratory are listed in SOP 1.

2.7.2 Quality control

Each central laboratory should prepare commutable specimens for QC monitoring as described in SOP 2.4. The appropriate batch size of the assay should be decided according to local requirements.

2.7.3 Standardization of the assay

For the standardized analytes, internationally qualified standards or CRMs should be measured to ensure traceability of test results.

2.7.4 Cross-comparison of values

A panel of sera prepared from healthy individuals should be used for cross comparison of values among the central labs based on linear regression analysis (reduced major major-axis regression). The magnitude of error in converting any RI from one assay platform to another can be estimated either by CV of slope, CV(b), or by SDR [the ratio of the standard error of the converted lower and upper limits, SE(LL) and SE(UL), relative to the SD comprising the converted RI, equal to SDR_{LL} and SDR_{UL}]. The optimal limit for CV(b) is 5.5% and that for SDR is 0.125 [18].

2.8 Cross-check testing between the central laboratory and each participating laboratory

To share the RIs derived for nonstandardized (non-harmonized) analytes such as hormones and tumor markers, cross-check testing should be conducted by asking each participating laboratory to retain one or two aliquots of serum from each volunteer and to measure them at the time of the centralized measurements. The number of samples required for the cross-check depends on the accuracy specification set by each laboratory. The error of conversion depends upon the correlation coefficient r and the data size n. From our previous studies, n≥20 allows conversion within a desirable level of error [14, 18] for the majority of the commonly measured analytes. The samples should be divided into multiple parts (at least four) and measured on different days to reduce the effect of between-day variation on the conversion. The linear structural relationship (reduced major axis regression) will be used to convert RIs established by the centralized assay to the values of each

participating laboratory [14, 18]. The required amount of additional blood depends upon the number of analytes each laboratory needs to measure locally for acquiring RIs through conversion based on the cross-check results.

2.9 Ethics

2.9.1 Declaration of Helsinki

The investigator should ensure that this study is conducted in full conformity with the current revision of the Declaration of Helsinki.

2.9.2 Ethics committee approval

The protocol, informed consent form, participant information sheet (questionnaire), and any proposed advertising material should be submitted to the ethics committee of each collaborator's institution for written approval. The investigator should also submit and obtain approval for all substantial amendments to the original approved documents.

2.9.3 Informed consent

See SOP 1.

2.9.4 Participant confidentiality

The investigating team should ensure that the participants' anonymity is maintained. The participants must be identified only by initials and an ID number on any electronic database sent outside the collecting institution. All documents should be stored securely and only accessible to the investigating team. The study should comply with the Data Protection Act, which requires data to be made anonymous as soon as is practical.

2.10 Data analysis and reports of the results

Data analysis should be performed according to SOP 3.

2.10.1 Validation of data

Those volunteers with overtly abnormal results should be secondarily excluded (e.g., hepatic or renal disease).

2.10.2 Analyses of source of variation of test results

Multiple regression analysis (MRA) should be performed analyte by analyte to identify factors related to variation of the test results [22]. The factors to be included in MRA are gender, age, BMI, smoking status, level of alcohol consumption, frequency and strength of physical exercise, dietary status, and, if available, ethnic group, the ABO blood group.

In the analysis, after combining all data across countries, dummy variables for the country or ethnic groups should be introduced to reveal regionality/ethnicity dependency. The dummy variables are also important to conduct regionality/ethnicity-adjusted analysis of sources of variation for the test results.

2.10.3 Partitioning criteria

The components of SD, i.e., between-country SD (SDcntr), between-sex SD (SDsex), between-age SD (SDage), and net between-individual SD (SDind) with removal of the component of within-individual variation, will be computed by 3-level nested ANOVA. The presence of significant regionality in the test results can be determined by taking a ratio (SDR) of SDcntr over SDindiv; an SDR >0.3 signifies the evidence of significant regionality in the test results [22]. However, its validity must be confirmed after adjusting for other possible confounding factors using MRA or similar procedures. SDR can be also computed for between-sex SD and between-age SD. The cut-off value of SDR=0.3 corresponds approximately to the midpoint of desirable and minimal proportion of analytical bias to the between-individual SD of 0.25 and 0.375 times the between-individual SD [18]. However, it is only a guide to consider whether to partition reference values according to that factor. It can be modified according to local need and policy in implementing the RIs. It is also necessary to note that, in dealing with regionality, if there is an isolated difference between only one or two countries with no differences among all other countries, ANOVA is insensitive for detecting regional differences. Therefore, it is recommended to apply an ad hoc analysis using the Harris-Boyd method be performed to evaluate the magnitude of the difference between any two groups.

If there are no apparent regional differences in test results, globally applicable common RIs may be derived by specifying conditions used for the derivation. Otherwise, reference values should be partitioned to derive region-, ethnicity-, or country-specific RIs.

2.10.4 Derivation of RI

The parametric method will be used after normalizing data based on a modified Box-Cox power transformation method [22] for derivation of RIs. Secondary exclusion of inappropriate subjects should be made to remove those with abnormal results attributable to highly prevalent conditions such as the metabolic syndrome, alcohol-related hepatic dysfunction, or diabetes mellitus. A multivariate iterative method called latent abnormal value exclusion (LAVE) [22] can be applied at the time for computing RIs as a method for tertiary exclusion.

When an analyte is not considered well-standardized, the RI derived by the centralized measurement may be converted to each laboratory's value using the linear regression parameters derived from the cross-check test results if the scatter around the regression line is within the allowable limit [CV(b)≤11%] [18].

3 Standard operating procedures

3.1 SOPs for recruitment (SOP 1)

3.1.1 Invitation to the study

It is advised to advertise the study by posters displayed in the wards and out-patient areas and by electronic invitations sent to staff members within the participating healthcare institution. It is also advisable to hold meetings to explain the clinical and scientific implications of the study and possible benefits for the laboratory and volunteers to obtain cooperation.

Give each volunteer the following:
1. An invitation to the study.
2. An explanation of the study (including/exclusion criteria).
3. A consent form (written in the local language, in accordance with the guidelines of the local ethics committee).
4. Procedures for participation.

3.1.2 Informed consent

During the introductory period, written and verbal information should be presented to the participants detailing the exact nature of the study and any risks involved in taking part. It should be clearly stated that the participant is free to withdraw from the study at any time for any reason with no obligation to give the reason for withdrawal. The participant must be allowed as much time as desired to consider the information and allowed an opportunity to question the investigator, their physicians or other independent parties to help decide whether they will participate in the study. Written informed consent should then be obtained, including the personal signature and signature date for both the participant and the person who presented and obtained the informed consent. The original signed form should be retained by the local representative.

3.1.3 Tabulation of volunteers by age and gender

In a practical flow of recruitment, one who agrees to participate in the study by reading the invitation is expected to contact the local representative. A balanced distribution of gender and age should be ensured by tabulating volunteers as shown below (Table 1). All participants must be older that 18 years. The main target range of ages is 18 to 65 years, for which an even distribution of age and gender is of the utmost importance to have comparability of test results across regions.

Individuals older that 65 years are also sought, as long as they match the inclusion criteria. Approximately 20% of the total number should be in this age group. The primary objective of including this age range is to evaluate age-related changes in test results, and thus, there is no need to balance the gender distribution of this group.

3.1.4 Appointment and preparation for blood sampling

It is advisable to make the appointment for the drawing of blood after volunteers for most of the subgroups in the above table have been recruited. Make appointments between 7:00 and 10:00 AM with 30-min intervals (i.e., 7:00, 7:30,…, 9:30, 10:00).

3.1.4.1 Sampling schedule
Prepare a sampling schedule as shown below, setting the date and time of sampling together with the volunteer's name and ID number. The ID number will be generated as AA-LL-###, with AA representing the country code; LL, the laboratory code within the country; and ###, the sequence number in the laboratory. The last 2 segments correspond to the same as the codes in Table 1.

	18–29 years	30–39 years	40–49 years	50–64 years	65–years
Male	01 001–01 050	01 101–01 150	01 201–01 250	01 301–01 350	01 401–01 450
	02 001–02 050	02 101–02 150	02 201–02 250	02 301–02 350	02 401–02 450
	03 001–03 050	03 101–03 150	03 201–03 250	03 301–03 350	03 401–03 450
Female	01 051–01 100	01 151–01 200	01 251–01 300	01 351–01 400	01 451–01 500
	02 051–02 100	02 151–02 200	02 251–02 300	02 351–02 400	02 451–02 500
	03 051–03 100	03 151–03 200	03 251–03 300	03 351–03 400	03 451–03 500

Table 1 Tabulation of volunteers by gender and age.
The numbers indicate the ID codes. ## ### (ID code) is shown as ## (laboratory code) + ### (donor code).

3.1.4.2 Reminders for volunteers in preparation for sampling

1. Make an appointment for blood collection when the volunteer agrees to participate.
2. Remind each volunteer of the following requirements before sampling:
   - Avoid unusual strenuous exercise for 3 days before the sampling.
   - Avoid sampling on the day after working a night shift.
   - Fast overnight (at least 10 h before the sampling).
   - Avoid excessive eating and/or alcohol intake the night before the sampling.
   - Avoid smoking just before the blood collection.

3.1.5 Procedures on the day of blood collection

3.1.5.1 Collection of questionnaire

Give the questionnaire (Appendices A and B, with the addition of any relevant local questions regarding demographic factors and lifestyle) at the time of phlebotomy. Participants will have their height, weight, and abdominal circumference measured by the local coordinators who also may help them complete the questionnaire. The ID labels should be pasted on both the questionnaire and consent forms. Make a photocopy of the questionnaire. The participants' consent form and the photocopy of the questionnaire will be kept confidential by the local representative of the study.

3.1.6 Procedures just before blood collection

3.1.6.1 Preparation of equipment for sampling and storage

The type of blood collection tubes, either plain or gel-separator tubes, must be determined in each country/laboratory; it is preferable to use the same sampling equipment as is used for routine testing. Assays for the most commonly measured analytes are not influenced by the tube type, except for some drugs or hydrophobic analytes [19–21]. However, it is recommended that each central laboratory investigates the possible differences in test results between the plain and gel-separator tubes by comparative measurements of several specimens. The type of tube being used must be recorded.

ID labels should be pasted onto vacuum blood collection tubes and storage containers. Make sure that the IDs match with those on the corresponding questionnaire and consent form for each participant.

3.1.6.2 Preparation of volunteers immediately before sampling

Volunteers should rest in a sitting position at least for 30 min before drawing blood. Hasty sampling after a volunteer rushes in causes stress-induced (inorganic phosphate, glucose, etc.) and postural changes (almost all proteins) in test results.

Smoking cigarettes just before blood collection is not allowed because smoking is known to affect the values for some enzymes (LDH and amylase) and glucose [23].

3.1.7 Procedures for drawing blood

- Apply the tourniquet 7–10 cm above the venipuncture site. The pressure should be set below the diastolic blood pressure for the smooth pooling of blood in the periphery. Never leave the tourniquet on for longer than 1 min.
- Do not clench the fist while drawing blood; this causes false elevation of serum potassium.
- Draw the volume of blood required.
- Invert each tube 180° (upside down) at least five times (in the presence of a clot activator in the sampling tubes).
- If the blood draw is interrupted before a tube is completely filled, the remaining vacuum should be removed by filling air to avoid vacuum-induced hemolysis [23].

3.1.8 Preparation of the serum and its aliquots

- Do not place the blood-filled tubes in direct sunlight or in low temperatures.
- At 30–60 min after sampling, centrifuge the specimen at 1200 g for 10 min at room temperature.
- Be sure the ID label on the blood collection tubes matches the ID labels on the sample containers (2 mL each). Transfer the serum into each of the containers.

3.1.9 Storage and shipment of the specimens

The containers with the serum aliquots (1–2 mL) should be well sealed and stored at −80°C within 2 h. All aliquots that are not being retained for local use will be shipped on dry ice to the central laboratory for collective measurement.

3.1.10 Thawing of the specimens and preparation of the samples before measurement

On the day of analysis, the samples must be thawed by letting them stand at room temperature for at least 1 h, avoiding direct sunlight because of its effect on bilirubin concentrations. Homogenization is then achieved by inverting the samples 10 times. Analysis must be performed within 4 h from the start of thawing.

3.1.11 Measurements for the cross-check survey

3.1.11.1 Procedure for cross-check testing

To share the RIs derived for nonstandardized analytes such as protein hormones and tumor markers, cross-check testing should be conducted by asking each participating laboratory to retain one or two aliquots of serum each from a part or all of the volunteers and to measure them locally near the time of the centralized assay for cross-comparison of the results.

3.1.11.2 The number of samples for cross-check testing

The number of samples required for the cross-check depends on the accuracy specification set by each laboratory. The recommended number is ≥10 for the standardized analytes and ≥20 for the nonstandardized analytes.

Procedures for the cross-check sample collection, and storage will be the same as previously stated in Protocol Section 8. Cross-check testing samples should be selected randomly. In addition, the samples for the cross-check should be divided into multiple parts (at least four) and measured on different days in order to reduce the effect of between-day errors on the conversion.

3.2 SOPs for analyses (SOP 2)

3.2.1 Sample size

The target sample size from each country should be at least 500 so that country-specific RIs can be derived in a reproducible manner. This number is sufficient for between-country comparison of test results. If there is an additional interest in exploring regional variations within a country, it is recommended that at least 120 (men and women, 60×2) samples from each local area be assayed to acquire sufficient power to test for between area differences.

3.2.2 Target analytes

The standardized analytes to be measured in common and to be compared worldwide should be decided and listed. Additional analytes to be measured are determined in each country based on clinical need and research interest. At the time of combined analyses of reference values, comparison will be made between results of all analytes that are available for comparison.

3.2.3 Central laboratory

One or two laboratories in each country should be chosen as a central laboratory that will provide collective measurements. The requirements for the central laboratory should be specified so that it (1) implements reliable measures for both short- and long-term quality control, (2) ensures traceability of test results for the standardized analytes based on CRMs and value-assigned sera for enzymes, and (3) participates in the alignment of test results across central laboratories, using the reference panel of sera.

3.2.4 Quality control

In addition to standard quality control materials supplied by manufacturers, each laboratory will prepare multiple commutable specimens (mini-panel) for QC monitoring. The desirable limits of bias should be specified beforehand. The desirable limits for between- and within-day CV are set as 1/2 of CV_I (within-individual CV listed in the Westgard website: www.westgard.com/biodatabase1.htm).

3.2.5 Standardization of assays

To ensure essential traceability of test results, internationally qualified standards or CRMs should be measured for the standardized analytes.

3.2.6 Cross-comparison of values

A panel of sera composed of 40 sera prepared from healthy individuals should be divided into at least four aliquots each, to be tested on different days in order to determine between-day variation of the test results. The between-day variation of test results should also be monitored by the QC specimens and recorded.

3.3 SOPs data analysis and report of the test results (SOP 3)

3.3.1 Data analysis

3.3.1.1 Analyses of source of variation of test results

MRA should be performed analyte by analyte to identify factors closely related to the test results [22]. The possible factors include gender, age, BMI, smoking status, level of alcohol consumption, frequency and strength of physical exercise, dietary status, ethnic group and, if available, the ABO blood group of the participant.

3.3.1.2 Partitioning criteria

The magnitude of between-country SD (SDcntr), between-sex SD (SDsex), between-age SD (SDage), and net between-individual SD (SDindiv) should be computed by 3-level nested ANOVA. Ratios of SDcntr over SDindiv >0.3 can be used as a guide to judge the presence of significant regionality in test results [22]. However, implication of regional differences should be evaluated by multiple regression analysis as described in the Protocol section.

3.3.1.3 Derivation of RI

For the derivation of RIs, the parametric method will be used after excluding samples with abnormal test results (secondary exclusion, as discussed above) and then normalizing the data by power transformation using the modified Box-Cox formula [22]. For tertiary exclusion, a multivariate iterative method called LAVE may be applied at the time of computing RIs [10, 11, 13, 14, 22].

3.3.1.4 Cross-comparison of values

A panel of sera composed of 40 sera prepared from healthy individuals should be divided into at least four parts and tested on different days to avoid bias attributable to between-day variation of the test results [18].

3.3.2 Report of test results

Only after all the specimens have been analyzed and the RIs have been derived should the test results be returned to the volunteers. A sheet with a given participant's printed test results should be sealed in an envelope with the ID label on the outside. The local representative should be asked to give this to the corresponding individual by referring to the name from the informed consent form with the same ID number. Each individual's test results and the newly-derived, country-specific RIs will be reported to the volunteers after all the measurements in each country have been completed, together with an explanatory sheet for the interpretation of the test results.

3.3.2.1 Report of cross-check test results and RIs

The cross-check the test results between the central laboratory and each of the local laboratories are analyzed by reduced major axis regression.

When an analyte is not considered well-standardized, judging from the regression line, the RI derived by the centralized measurement can be converted to each laboratory's value using the linear regression slope and intercept [18].

4 Discussion and conclusions

Despite the long-recognized importance of RIs in the clinical decision-making process, the implementation of RIs in most clinical laboratories is still incomplete [24]. The derivation of RIs on a national level by conducting a multicenter study that follows a common protocol and comprehensible SOPs, and the secondary integration of the results at a global scale, is probably the most effective way to seek globally applicable common RIs [25]. The success of globalization depends upon the absence of regional and/or ethnic differences in test results, whereas their presence provides invaluable information regarding those differences. To produce scientifically valid results that can be merged to make comparisons across the world, multicenter studies in each country should be conducted utilizing a common protocol and SOPs.

This study aimed to produce a protocol and SOPs that can be used in harmony by any laboratory anywhere in the world, thus ensuring repeatability and transferability

of reference values. To achieve this, the CSLI/IFCC C28-A3 guideline, entitled "Defining, Establishing and Verifying Reference Intervals in the Clinical Laboratory" was used as the basis but was expanded to provide the necessary steps in more detail so that every participating laboratory follows exactly the same SOPs in the common protocol. The current study was planned specifically as a multicenter, multicountry study, and a common protocol was essential.

There are several important differences between the current protocol and the C28-A3 document. Modifications were made to C28-A3 regarding the inclusion/exclusion criteria in order to show more clearly how healthy reference individual were defined during selection, although these criteria are essentially pragmatic and there may be subtle differences in interpretation of what constitutes healthiness by different cultures and investigators. Appendices A, B, and C were deemed necessary to achieve the required level of standardization. Ethnic origins are questioned in detail in Appendix A. Additional items were included in the questionnaire (Appendix B) to obtain more quantitative information regarding alcohol consumption, physical activities, menstrual cycle, and medications to see how these factors influence test values. Appendix C lists the target analytes to be measured in common by the participating laboratories.

In the new protocol, standardization procedures are described in detail so that all central laboratories are made aware of the necessity to comply with them in full. This protocol also differs from previous guidelines with regard to the required sample size [i.e., at least 500 reference individuals for each country instead of 240 volunteers (120 of each gender)], calculation methods for RIs, and partitioning criteria to be used [22]. In addition, the protocol requires this study to be performed using a panel of sera for harmonizing results across central laboratories. A major difference between this protocol and SOPs and the previous C28-A3 guidelines is that, in order to enable comparison of test results between the central laboratory and each of the local laboratories, the cross-check testing procedure is also described in detail at each stage [14, 18], thus allowing comparison and transfer of the centrally derived RI to the values of local laboratories.

The protocol and the SOPs described here give an in-depth coverage of the scientific criteria and requirements to ensure valid results. They can be used, modified, and/or adopted for similar multicenter studies to establish common RIs. The appropriateness of the protocol and the SOPs described here will be evaluated through their implementation in the currently ongoing worldwide multicenter study on reference values.

Conflict of interest statement

Appendix A

A: White	
British	
Irish	
Any other White background	(please write in)
B: Mixed	
White and Black Caribbean	
White and Black African	
White and Asian	
Any other mixed background	(please write in)
C: Asian or Asian British	
Indian	
Pakistani	
Bangladeshi	
Any other Asian background	(please write in)
D: Black or Black British	
Caribbean	
African	
Any other Black background	(please write in)
E: Chinese or other ethnic group	
Chinese	
Any other (please write in)	
Not stated	
Not stated	
F: For US	
Hispanic or Latino	
American Indian or Alaska Native	
Native Hawaiian or other Pacific Islander	
Black/African American	

Appendix Table 1 Ethnicity-monitoring questionnaire.
Citation from Office for National Statistics: http://www.ons.gov.uk/ons/guide-method/measuring-equality/equality/ethnic-nat-identity-religion/ethnic-group/index.html#8 (accessed 23 April 2013).

Appendix B

INCLUSION CRITERIA

1. Do you consider yourself healthy?   ☐ YES   ☐ NO
2. Are you at least 18 years of age?   ☐ YES   ☐ NO

EXCLUSION CRITERIA

1. Do you have diabetes and are treated with oral therapy or insulin?   ☐ NO   ☐ YES
2. Do you have or have you had chronic liver or kidney disease?   ☐ NO   ☐ YES
3. Do you have results from a blood test that point to a severe disease?   ☐ NO   ☐ YES
4. Have you been hospitalized or seriously ill in the past 4 weeks?   ☐ NO   ☐ YES
5. Have you given blood as a donor in the previous 3 months?   ☐ NO   ☐ YES
6. Are you a known carrier of HBV, HCV, or HIV?   ☐ NO   ☐ YES
7. Are you pregnant or within one year after childbirth?   ☐ NO   ☐ YES
8. Have you participated in a research study involving an investigational drug in the past 12 weeks?   ☐ NO   ☐ YES

DEMOGRAPHICS

Are you fasting ?   ☐ YES   ☐ NO   Last food intake: ☐ (AM / PM)

Age: ☐   Gender: ☐ Male ☐ Female   Today's date: ☐
Time of collection: ☐

Body Length: ☐ cm   Body Weight: ☐ kg   Abdominal Circumference: ☐ cm

ABO blood type: ☐ A   ☐ B   ☐ AB   ☐ O   ☐ ? not known

HEALTH STATUS AND MEDICAL HISTORY

1: Are you vegetarian (no meat or fish):   ☐ NO   ☐ YES
2: Do you eat a special diet?   ☐ NO   ☐ YES
−−If yes, what?

3: Have you been sick within the past 4 weeks?   ☐ NO   ☐ YES
−−if yes, describe the illness and when.

4: Have you been hospitalized in the last 6 months?   ☐ NO   ☐ YES
−−If yes, what for and when?

5: Do you have any allergic condition (pollinosis, atopic dermatitis, asthma, etc) ?   ☐ NO   ☐ YES
−−If yes, what? (describe the condition, is it currently active?)

6: Do you have high blood pressure?   ☐ NO   ☐ YES

7: Are you currently under a doctor's care?   ☐ NO   ☐ YES
−−If yes, what for?

8: Are you taking any prescribed medication on a regular basis?   ☐ NO   ☐ YES
     (including diet pills, anti-hypertensive, antacids, or allergy medicine, etc.)
−−If yes, what? (describe the name, dose, and frequency for each medicine.)

General health screening questionnaire

9: Do you take vitamin supplements or herbal remedies? ☐ NO   ☐ YES
−−If yes, what? (describe the name, dose, and frequency for each medicine.)

10: Have you taken any pain relievers in the past 4 weeks? ☐ NO   ☐ YES

11: Are you exposed to any hazardous chemicals in your job? ☐ NO   ☐ YES
−−If yes, what?

WOMEN

1: When was your last period? ☐

2: How is your menstrual cycle? ☐ regular   ☐ irregular   ☐ menopause   ☐ under hormone therapy
or taking contraceptive pills

−−If "**regular**", what is the averge lengh of the cycle? ☐ days/

ALCOHOL

1: Have you had a drink in the last 48 hours? ☐ NO   ☐ YES
−−If yes, what?

2: How much do you drink in a typical week? ☐ NONE

			∗ Calculation by the invetigator:
Beer	☐ litres		Amount of alcohol consumption
Cider	☐ litres		in grams of ethanol per week
Wine	☐ glasses (assume 6 glasses per standard sized bottle)		
Spirits	☐ bottles		☐ g/week

3: For how many months/years has this been typical? ☐ years

SMOKING

1: How many cigarettes do you smoke ? ☐ per day ☐ per week
2: If you roll your cigarettes, how many ounces per week do you smoke? ☐ ounce per week
3: If you smoke cigars and pipes, how many days per week do you smoke? ☐ cigar or pipes/week
4: How many years have you smoked? ☐ years
5: If you quit smoking within a year, please describe when you quit, and how much you smoked (amount and years).

PHYSICAL ACTIVITIES AND EXERCISE

1. In a usual day, how many hours do you stand: (1) ☐ hours/day (include exercise, commuting)
   In a usual day, how many hours do you sit: (2) ☐ hours/day
   Note: 24 hours = (1) + (2) + (hours of sleep or other lying posture/day).

2. Do you exercise regularly? ☐ NO   ☐ YES

 −−If yes, how many days per week? ☐ days/week

   Please describe your regular exercise (name or type of exercise, duration/day).

3. How many days in the past week have you performed physical activities during your work or exercise when your heart has beat
   faster and your breathing was harder than normal for a total of 30 minutes or more per day? ☐ days/week
4. How many days in a typical week have you performed such strenuous activity? ☐ days/week

Appendix C

Target analytes

The analytes below are to be measured in common. However, each country may add or omit any analytes if required. At the time of combined analyses of reference values across the world, comparison will be made between results of whatever analytes are available for comparison.

Chemical analyses

The following analytes constitute "standard analytes":
– Enzymes: AST, ALT, ALP, LD, GGT, CK, and amylase
– Electrolytes: sodium, potassium, chloride, calcium, inorganic phosphate, iron, and magnesium
– Miscellaneous: total protein, albumin, creatinine, urea, uric acid, total bilirubin, and glucose
– Lipids: triglycerides, total cholesterol, HDL-C, and LDL-C (these analytes should be measured as part of the assessment of nutritional status).

Immunoturbidimetry

This includes CRP, IgG, IgA, IgM, C3, C4, transferrin, TTR (prealbumin), and cystatin C.

Immunoassays

It is preferred that the following nonstandardized but commonly measured analytes also be measured in common to allow international comparison of results:
– Tumor markers: ferritin, AFP, CEA, CA125, and PSA
– Endocrinology: TSH, prolactin, cortisol, and PTH
– Miscellaneous : vitamin B12 and folate

CA19-9, CA15-3, fT4, and fT3 have been excluded due to known method-related variation and failure of their test results to be made comparable based on previous cross-check testing. Those analytes that are very unstable or require a special sampling tube should also be excluded.

References

1. Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1986) on the theory of reference values. Part 1.The concepts of reference values. Clin Chim Acta 1987;165:111–8; J Clin Chem Clin Biochem 1987;25:337–42.
2. Petit Clerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 2. Selection of individual for the production of reference value. Clin Chim Acta 1987;170:S1–12; J Clin Chem Clin Biochem 1987;25:337–42.
3. Solberg HE, Petit Clerc C. International Federation of Clinical Chemistry (IFCC). Approved recommendation on the theory of reference values. Part 3. Preparation of individuals and collection of the specimens for the production of reference values. Clin Chim Acta 1988;177:S3–11; J Clin Chem Clin Biochem 1988;26:593–8.
4. Solberg HE, Stamm D. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1991) on the theory of reference values. Part 4. Control of analytical variation in the production, transfer and application of reference values. Clin Chim Acta 1991;202:S5–11; J Clin Chem Clin Biochem 1991;29:531–5.
5. Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Clin Chim Acta 1987;170:S13–32; J Clin Chem Clin Biochem 1987;25:645–56.
6. Dybkaer R, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. Clin Chim Acta 1987;170:S33–42; J Clin Chem Clin Biochem 1987;25:657–62.
7. Rustad P, Felding P, Lahti A, Petersen PH. The Nordic Reference Interval Project 2000. Scand J Clin Lab Invest 2004;64:271–84.
8. Fuentes-Arderiu X, Ferrer-Masferrer M, Gonzales-Alba JM, Villarino-Gonzales MI, Arrimadas-Esteban E, Cabrero-Olive D, et al. Multicentric reference values for some quantities measured with the Elecsys 2010® analyser. Clin Chim Acta 2001;304:143–6.
9. Jones GR, Barker A, Tate J, Lim C-F, Robertson K. The case for common reference intervals. Clin Biochem Rev 2004;25:99–104.
10. Ichihara K, Itoh Y, Min WK, Yap SF, Lam CW, Kong XT, et al.; Committee on Plasma Proteins, IFCC. Diagnostic and epidemiological implications of regional differences in serum concentrations of proteins observed in six Asian cities. Clin Chem Lab Med 2004;42:800–9.
11. Ichihara K, Itoh Y, Lam CW, Poon PM, Kim JH, Kyono H, et al. Sources of variation of commonly measured serum analytes among 6 Asian cities and consideration of common reference intervals. Clin Chem 2008;54:356–65.
12. Ceriotti F, Henny J, Queraltó J, Ziyu S, Özarda Y, Chen B, et al.; IFCC Committee on Reference Intervals and Decision Limits (C-RIDL); Committee on Reference Systems for Enzymes (C-RSE). Common reference intervals for aspartate aminotransferase (AST), alanine aminotransferase (ALT) and γ-glutamyl transferase (GGT) in serum: results from an IFCC multicenter study. Clin Chem Lab Med 2010;48:1593–601.

13. Ichihara K, Ceriotti F, Tam TH, Sueyoshi S, Poon PM, Thong ML, et al. The Asian project for collaborative derivation of reference intervals: (1) strategy and major results of standardized analytes. Clin Chem Lab Med 2013. DOI: 10.1515/cclm-2012-0421 [Epub ahead of print].

14. Ichihara K, Ceriotti F, Mori K, Huang YY, Huang YY, Shimizu Y, et al. The Asian project for collaborative derivation of reference intervals: (2) results of non-standardized analytes and transference of reference intervals to the participating laboratories on the basis of cross-comparison of test results. Clin Chem Lab Med Med 2013. DOI: 10.1515/cclm-2012-0421 [Epub ahead of print].

15. CLSI and IFCC, C28-A3 document. Defining, establishing and verifying reference intervals in the clinical laboratory: approved guideline, 3rd ed. Vol. 28, 2008:1–76.

16. Rustad P, Felding P, Lahti A. Proposal for guidelines to establish common biological reference intervals in large geographical areas for biochemical areas for biochemical quantities frequently in serum and plasma. Clin Chem Lab Med 2004;42:783–91.

17. Fuentes-Arderiu X, Mas-Serra R, Aluma-Trullas A, Marti-Marcet MI, Dot-Bach D. Guideline for the production of multicentre physiological reference vales using the same measurement system. A proposal of Catalan Association for Clinical Laboratory Sciences. Clin Chem Lab Med 2004;42:778–82.

18. Ichihara K, Ozarda Y, Klee G, Straseski J, Baumann N, Ishikura K. Utility of a panel of sera for the alignment of test results in the worldwide multicenter study on reference values. Clin Chem Lab Med 2013;51:1007–25.

19. Comparison of plastic versus glass evacuated serum-separator (SST) blood drawing tubes for common chemistry determinations. Clin Chem 1992;38:1474–8.

20. Bush VJ, Janu MR, Bathur F, Wells A, Dasgupta A. Comparison of BD Vacutainer SSTTM plus tubes with BD SSTTM II plus tubes for common analytes. Clin Chim Acta 2001;306:139–43.

21. Stankovica AK, Parmar G. Assay interferences from blood collection tubes: a cautionary note. Clin Chem 2006;52:1627–8.

22. Ichihara K, Boyd J. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537–51.

23. Tamechika Y, Iwatani Y, Tohyama K, Ichihara K. Insufficient filling of vacuum tubes as a cause of microhemolysis and elevated serum lactate dehydrogenase levels. Use of a data-mining technique in evaluation of questionable laboratory test results. Clin Chem Lab Med 2006;44:657–61.

24. Ceriotti F. Prerequisites for use of common reference intervals. Clin Biochem Rev 2007;28:115–21.

25. Ceriotti F. Common reference intervals: the IFCC position. Clin Biochem 2009;42:297.