Medical laboratory findings require a measured result accompanied by a reference interval. The reliability of both values (measured result and reference limit) should be based on the same analytical reliability criteria (e.g. the same bias), also the biological variables of the reference subgroup selected should match the biological variables of the population served by the laboratory. If this is not the case, the diagnostic efficiency is reduced. Thus, the diagnostic non-specificity as a measure for diagnostic efficiency (indicated by an increase of false positive results) is decreased if, e.g. the upper reference limit of the reference subpopulation is higher than the limit of the population to be served. The biological variables most often studied are gender and age.

Two divergent approaches are presently applied for estimating reference limits, direct and indirect methods. The crucial difference between direct and indirect methods is that direct methods select particular individuals because of their properties. These individuals form a reference subpopulation. Most indirect approaches [1–3] are based on separating the combined data set of diseased and non-diseased individuals (obtained by data mining) into at least two distributions (one distribution of “healthy” individuals and one distribution of diseased patients) by resolution techniques. Thus, indirect methods select a reference subpopulation in which properties of individuals are not identified [4].

Indirect approaches for the estimation of reference limits usually apply data mining from the population served by the laboratory, whereas direct models (although the present “gold standard”) often derive their limits from a group of young, probably “healthy” people. Both approaches have their benefits [5, 6]. The advantages of the indirect procedures have been mentioned by many authors [e.g. 4–6]: higher numbers of reference individuals available with less expenses allowing better partitioning and higher diagnostic efficiency, no ethical problems, as well as comparable pre-examination procedures, age distribution, population characteristics and biological variables, etc.

With data mining, the disease prevalence in the reference subpopulation can be reduced by exclusion criteria. These criteria can be extended in such a way that only a small group of highly selected individuals remains which may be considered as probably free of the disease considered (non-diseased), as e.g. free of acute myocardial infarction (AMI) in the case of high-sensitive cardiac troponin T (hs-cTnT).

Monneret et al. [7] re-evaluated the influence of gender and age on the reference limits of hs-cTnT by applying a data mining approach. They defined a central distribution of AMI free patients by excluding most patients with probable AMI, and reduced the number of patients by strict exclusion/selection criteria until they obtained a relatively small reference group of patients which are probably free of AMI. Individuals declared as “free of AMI” were identified by their dynamic of the hs-cTnT values measured. This is an interesting alternative approach which should be classified as direct. However, it may be a semantic question whether such a technique is called indirect or direct with a-posteriori selection of a non-diseased subpopulation.

While the approach of Monneret et al. [7] is innovative and relevant, there are some critical issues. One major point is the treatment of so-called outliers. In clinical chemistry, outliers are somehow erroneous results. In the present context, outliers mean results supposed not to belong to the reference group which is characterized by the probable absence of AMI. The authors excluded data by a combination of clinical and statistical aspects. Thus, changes in serial hs-cTnT concentrations were used as an exclusion criterion for a selected group of “non-AMI”. But Hammarsten et al. [8] reported that changes >20% were unexpectedly common in hospitalized patients. The technique of Monneret et al. [7] led to a reduction from about 170,000 patients’ files to 1548 troponin values for men and 1159 values for women, numbers which may be considered as insufficient for
further partitioning in the case of a very skewed distribution [9]. Furthermore, tolerance limits (instead of confidence limits) are missing, artificial jumps occur between the age groups selected, the proposals are only based on imprecision neglecting the influence of bias (especially between different manufacturers), and the stability of the analytical procedure during the 3.8 years selection period is not indicated. Despite these limitations, the key message of an age-dependent upper reference limit appears relevant.

Considering several age groups instead of an unstratified age group (e.g. from 18 to 99 years), increases the diagnostic efficiency. The authors proposed three age groups with a step-like increase of the troponin T concentration. If the age switches from, e.g. 50 to 51, a 1-day difference of the upper reference limit suddenly occurs with men from 19 to 33 ng/L troponin T which is unrealistic. In most cases, age dependency does not occur in “jumps” between several population groups, but rather follows a continuous kinetic as has been demonstrated by Zierk et al. [10, 11].

Step-like patterns may be acceptable as long as a continuous function is not available. The hierarchy for establishing age-dependent reference limits should be: unstratified if not enough data are available, step-like approach, continuous approach. Age-related RL can be derived for each individual result if a continuous function has been established. Age-dependency of reference limits is more often observed than generally known. Its study should be included in software programs for estimating reference limits as, e.g. in the package available on the home page of the German Society for Clinical Chemistry and Laboratory Medicine (DGKL) [12]. Proposals to check the relevance of differences between two reference limits are available [13, 14].

Partitioning with several age-groups, and especially the establishment of continuous reference bands requires large datasets which usually cannot be obtained by direct approaches with acceptable expenses. Therefore, Monneret et al. [7] applied a data mining approach. Data mining of hospitalized inpatients vs. “healthy” individuals causes still many debates in the literature. This is mainly due to numerous biological variables as, e.g. age, gender, posture, diurnal variations, etc. which are not considered. Therefore, some authors prefer outpatients for data mining (e.g. private laboratories serving mainly general practitioners). In this case, the derived reference limits are closer to those determined by direct approaches mostly using young “healthy” people walking around during daily work and with sample taking during the morning.

But with troponin, outpatients are probably less suited for indirect approaches. About 30% of the troponin values may be below the detection limit. Kairisto et al. [15] argued that patients hospitalized with chest pain but later proved to not have suffered from myocardial infarction would be ideal reference subjects for laboratory markers of myocardial infarction.

Whereas the age dependency appears convincingly demonstrated for troponin T by Monneret et al. [7], the influence of sex remains debatable. Earlier studies reported a clear sex difference with higher values in men than in women [16–18]. FDA approval of the hs-cTnT assay in the United States specified sex-specific cut-off values as 14 ng/L for women and 22 ng/L for men [19]. Monneret et al. [7] also showed slightly higher values in men which they considered as clinically not relevant. But in their table 2, they demonstrated that confidence intervals did not overlap indicating that a partitioning for gender could be justified on statistical grounds. Age stratification appears particularly interesting in the elder group above 71 years. In this age group women had higher concentrations than men contrary to the younger age groups. The authors considered this difference as uninterpretable due to the different age-distribution. However, this group should be reinvestigated with a larger data set and smaller age intervals. A continuous approach as mentioned above perhaps could lead to a clearer situation.

Partitioning for age and gender, and for even more variables as, e.g. diurnal, seasonal, regional variations, increase the number of reference limits. In extreme situations, most measured results may require their own reference limit, which may confuse interpretation of laboratory results in daily practice. This can only be handled if partitioning is integrated in laboratory information systems. Then, the number of reference limits becomes irrelevant and also small partitioning effects should be considered (as, e.g. gender differences of troponin T values).

The reference limits of many measurands are influenced by diurnal variation. This biological variable is often neglected if reference limits derived by direct methods are compared with limits estimated by indirect approaches if different sampling times are applied. Klinkenberg et al. [20] observed a diurnal rhythm of troponin T. Troponin T has its peak during the morning and its trough during the afternoon. This biological variable was not considered by Monneret et al. [7].

Usually, reference limits are related to a “healthy” subpopulation. If the term reference limit is applied to a group of hospitalized patients which most probably do not suffer from AMI, terms like cut-off values, decision limit, or more favorable action limits may be more appropriate.
As many other authors, Monneret et al. [7] shifted the upper reference limit of hs-cTnT from the 95% inner interval to the 99% interval to get a diagnostic decision point. The term reference interval should be restricted to the 95% interval. Diagnostic or therapeutic cut-off values should be termed action limits [21]. If the 99% limit is chosen as the action limit, it has a high transparency in comparison with less transparent limits as, e.g. the 200 mg/dL plasma cholesterol concentration or the 7 mmol/L blood glucose concentration.

However, regarding the 99% limit as a reference limit or action limit is a pure semantic problem which does not influence the key message of the publication of Monneret et al. [7] concerning the age-dependency of the upper reference, resp. action limit.

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

Research funding: None declared.

Employment or leadership: None declared.

Honorarium: None declared.

References


Prof. Rainer Haeckel, Institute for Laboratory Medicine, Katrepeler Landstr. 45 E, Bremen, 28357, Germany, E-mail: rainer.haeckel@t-online.de