Reference Intervals Developed from Data for Hospitalized Patients: Computerized Method Based on Combination of Laboratory and Diagnostic Data

Timo Kouri,1,2,6 Veli Kaiisto,1 Arja Virtanen,1 Esa Uusipaikka,7 Allan Rajamäki,4 Hanna Finneman,5 Kale Juva,4 Timo Koivula,2 and Veikko Näätä1

We utilized the databases of a hospital information system to select for determination of reference values various individual hospitalized patients on the basis of their diagnoses at discharge. The nonparametric 2.5–97.5% “health-related” reference intervals were calculated for hemoglobin concentration, mean corpuscular volume (MCV), and erythrocyte count for both sexes. After excluding patients with diseases possibly affecting erythrocyte variables, we obtained a final group of 1786 women and 1450 men, ages 20–65 years, who were studied in age groups of 20–30, 30–45, 45–55, and 55–65 years. The upper reference limits of the MCV results obtained from hospitalized patients were higher than those produced conventionally from healthy individuals, as would be intuitively suggested by clinical experience. This method, based on selection by diagnosis, could be applicable to various analytes measured in hospital laboratories, provided sufficient data are available as databases.

Indexing Terms: reference values/databases/hospital information systems/erythrocytes/hemoglobin

The best reference values for an individual are derived from her or his own prior data. Because these are not often available, population-based reference values have been widely used. Usually, reference intervals have been established from a population of healthy, nonpregnant, nonobese individuals, who have neither ingested any drugs nor smocked prior to sample collection (1, 2). However, the procedure for a priori selection of qualified reference individuals from representative populations is cumbersome; moreover, the most rigorous recommendations for exclusion criteria may result in only a minor fraction of healthy people being accepted (3). In addition, healthy ambulatory individuals are not optimal references for hospitalized patients, because of differences in, e.g., body posture, physical activity, diet, state of mind, or diurnal rhythm in the hospital compared with those prevailing in regular life.

When diagnostic classifications are being made, reference intervals are used as yardsticks to discern specific diseases from general disability (4). From that point of view, the best reference for a hospitalized patient is another patient not affected by the disease in question, but living under the same conditions as the patient whose laboratory result is being interpreted (5). Moreover, laboratory results from samples of reference individuals are ideally produced in the same way as those of the investigated patients, with respect to the analytical variation as well as the preanalytical process (including sample handling, transport, and storage). In addition, separate reference values should be collected for ambulatory and hospitalized patients because of the postural difference (sitting or recumbent position). In many cases, the availability of reference data from several populations—i.e., reference values for both control and diseased populations—would be useful in the evaluation of a single laboratory result (4, 6). In the recommendation of the International Federation of Clinical Chemistry (IFCC), several types of reference populations are accepted, depending on the intended use of the derived reference intervals, as long as the reference groups used are sufficiently characterized (1).

Previous estimations of reference intervals from hospitalized patients (e.g., 7–11) have involved deriving reference intervals either mathematically or graphically from unselected patient data, or using unselected data as such. Without knowledge of the number or nature of subpopulations among unselected patients, a meaningful interpretation of such a reference interval may be difficult (9). Indirect methods may widen the reference interval or shift it toward pathological values, depending on the proportion and the level of sickness-related values among the unselected data (12). That is why the indirect methods described previously are not recommended by IFCC (1, 5, 13).

The final discharge diagnoses collected into a hospital information system allow an a posteriori selection of relevant patients to be used as reference individuals. After combination with laboratory data, the results for selected patients may be used to estimate reference intervals for different analytes. In this way, both “healthy” and “diseased” reference populations can be delineated. Here we report details of such an approach, as applied to
selection of adult health-related reference values for three erythrocyte variables measured directly by automatic analyzers: hemoglobin (Hgb), mean corpuscular volume (MCV), and red blood cell (RBC) count.\(^7\) Preliminary data of this approach have been reported earlier (14).

**Subjects and Methods**

**Analytical Methods**

Blood specimens for hematological analyses were drawn into evacuated collection tubes containing EDTA (Vacutainer Tubes, no. 367652; Becton Dickinson, Lincoln Park, NJ) as part of a clinical investigation of the patients. Red cell parameters were analyzed by Coulter Counter S-Series (S Plus VI and T-880; Coulter Electronics, Hialeah, FL: \(\sim 90\%\) of the results) or Technicon H6000 analyzer (Miles, Tarrytown, NY: \(\sim 10\%\) of the results), calibrated to give the same concentration for the analytes. For internal quality control, stable control samples and leftover blood from patients or blood donors were used; in addition, individual results were inspected by medical laboratory technologists to detect random errors.

**Selection Methods**

*Generation of the combined database.* Turku University Hospital is a central hospital with 1200 beds, comprising all main clinical specialties. Pertinent diagnosis codes (the Finnish version of the WHO International Classification of Diseases, ICD-9, 15) are given by the responsible clinician for each treatment period, when the patient is discharged. From the discharge diagnosis database (programmed in M language on a VAX computer), the following information was obtained for each patient and treatment period: days of admission and discharge, principal diagnosis, and one to three possible additional diagnoses. The laboratory data contained 80,675 measurements of RBC variables. We extracted from the laboratory database (also in M language) the time of sample collection, the names of the laboratory tests, their results, and the clinical wards of the patients. The two databases were combined by using social security numbers, which were then removed (and replaced with random recoding) to protect the privacy of the patients. Only age and sex were retained in the combined database to make later partitioning possible (Fig. 1).

For a large number of outpatients, no diagnostic information was available in the database, because the treatment period was not completed yet. Therefore, this study was focused on inpatients. To avoid a possible bias from repeated measurements, we included only the first available test result for each patient in the study, thus forming a database of test results “on admission.” This made available measurements of RBC variables as well as diagnostic data for 27,400 patients. The subjects were further selected to be limited to adult patients (ages 20–65 years). The final group for analysis thus consisted of 12,879 inpatients, ages 20 to 65 years. These experimental procedures complied with the ethical standards of Helsinki Declaration of 1983.

*Selection by diagnosis.* The Finnish version of ICD-9 classification was considered carefully and diagnoses of diseases and conditions possibly affecting RBC variables were listed as exclusion criteria at the four-digit level of the classification (Table 1). The criteria were selected by an experienced hematologist (A.R.) on the basis of conceivable pathogenetic mechanisms. States and conditions affecting plasma volume (pregnancy-associated conditions, endocrine and metabolic diseases, and gastrointestinal, renal, cardiac, and similar diseases) were excluded, as were those influencing synthesis or degradation of Hgb or RBC (anemias, chronic pulmonary diseases, diseases of bone marrow, and chronic and many acute infections) and those associated with possible bleedings (traumas, malignant tumors, some benign tumors, diseases of coagulation, abnormal menstrual bleedings). The actual selection of patients was performed by a microcomputer using the SAS statistical program (SAS Institute, Cary, NC) (Fig. 1). The sizes of the reference groups thus obtained were 1967 women and 1613 men inpatients for Hgb measurements; slightly fewer were selected for MCV and RBC count.

---

\(^{7}\) Nonstandard abbreviations: Hgb, hemoglobin; MCV, mean corpuscular volume; RBC, red blood cell (erythrocyte); and ICD, International Classification of Diseases.
The diagnostic subgroups for these patients are shown in Table 2.

Statistical evaluation. Patients with the 12 most frequent accepted diagnoses were compared with the remaining diagnoses combined as a 13th subgroup, to determine possible differences in the concentrations of the results (Kruskal–Wallis analysis of variance). Detailed comparison of individual diagnosis subgroups was performed against the rest of each reference group to discover possible differences in the distributions (Kolmogorov–Smirnov test for two groups) or in the levels of concentrations of the results (Wilcoxon’s test for two groups). In this way, statistically important subgroups could be discerned and the selection principle reevaluated.

Reference Intervals
In the diagnosis-selected reference groups, some collection errors still had to be eliminated: Tests done during postoperative periods could be found by comparing laboratory data with stored dates and times of surgery in the anesthesia database; such results were eliminated. Results from the first week of the year were also excluded to remove the patients admitted before the yearshift. Consequently, the total number of patients was reduced to 1786 women and 1454 men (for Hgb results), and 1784 women and 1450 men (for MCV and RBC counts). Reference intervals were produced for Hgb, MCV, and RBC counts for both women and men by nonparametric methods, with use of the 2.5% and 97.5% reference limits with 95% binomial confidence limits (REFVAL program; designed by H. E. Solberg, Department of Clinical Chemistry, Rikshospitalet, Oslo, Norway).

Results
Selection. In terms of the ICD classification, patients with several hundred different diagnostic codes of diseases possibly affecting RBC parameters were excluded from the reference group (at the four-digit level of the classification, Fig. 1, Table 1), forming a total of 9338 excluded diagnoses among the women and 3763 diagnoses among the men (Table 2; as an example of coding errors, three men had “pregnancy-associated” diagnoses). The most frequent individual diseases excluded are listed in Table 3. Among the women, pregnancy-
associated conditions led most often to exclusion. Among the men, the exclusion group was dominated by traumas, cardiovascular diseases, and respiratory diseases.

**Statistical evaluation of selection.** Among the men, significant differences in the Hgb results were seen between the 13 subgroups classified by diagnosis (the 12 most-frequent diagnostic subgroups vs the remaining group “others”: $P = 0.0001$, Kruskal–Wallis analysis of variance); such was not the case for the 13 subgroups of women ($P = 0.31$; Table 4). The diagnoses associated with a difference in the Hgb concentrations of male patients were: sprained ankle [63 diagnoses ($P = 0.0001$, Kolmogorov–Smirnov test]; lower Hgb values than those of patients with other diagnoses], deformed nasal septum [41 diagnoses ($P = 0.043$); higher Hgb values], myocardial infarction [31 diagnoses ($P = 0.044$); lower Hgb values], and angina pectoris [92 diagnoses ($P = 0.056$); lower Hgb values]. Because hypoxia may predispose to cardiac events, we excluded these two cardiac diseases from the reference group in the statistically purified sample, although changes in Hgb concentration do not occur at the speed of usual cardiac attacks. However, ankle sprains and nasal deformations were accepted into the reference group, since we could not conceive a pathogenetic mechanism for lower or higher Hgb values of these patients. Thus, a total of 1813 diagnoses were left in the “healthy” male patient reference group (Table 2; Table 4). The average number of different diagnoses per individual was smaller in the diagnosis-selected group than in the excluded group, as expected.

Urogenital diseases dominated in the women’s reference group (e.g., vaginal prolapse, stress incontinency, sterility) (Table 4). The most frequent single subgroup included in the reference group were women admitted for sterilization without any disease. The men’s reference group was dominated by noninflammatory connective tissue diseases, such as sciatic syndrome and osteoarthrosis. Minor traumas or consequences of earlier traumas were the second largest subgroup. A remarkable number of neurological diseases also characterized the men’s reference group (Table 2).

**Reference intervals obtained.** After reevaluation of selection criteria, the groups for each sex were further refined to eliminate technical mistakes related to collection errors. For some patients, the first Hgb measurement recorded into the database was a postoperative value. Final distributions of Hgb, MCV, and RBC counts of the reference patients are shown together with those of the unselected inpatient groups, ages 20–65 years, in Figs. 2 and 3. To show dependency on age, we further divided the reference groups by age into four subgroups and estimated the reference limits with corresponding confidence intervals. The results were tabulated for
practical applicability (Table 5). Even the iteratively transformed distributions usually failed to approximate a gaussian shape (REFVAL program); therefore, we considered nonparametric estimates of reference intervals more reliable. The effect of menstrual bleeding was seen in the Hgb concentration and RBC counts of women of fertile age. Among the men, Hgb concentra-

tions and RBC counts of the younger subgroups were higher than those of the older subgroups.

Discussion
Hospitalized Reference Population
A hospitalized reference population can be seen as a formally characterizable reference category. One can argue that the clinical problem of differential diagnosis is not to discern "health" from "disease," but to discern a particular disease from other diseases. In that sense, other hospital patients may be better references than walking healthy individuals, who often are not even matched with the patient for age. Moreover, pertinent distributions of laboratory results from both health-related and diseased cases can be calculated only if the results have been obtained in a similar way for both groups. The prerequisites for extracting health-related reference intervals from hospitalized patients include the fact that a laboratory test is often ordered for patients who then are found to be unaffected with respect to the analyte being tested. This consideration may limit the applicability of this method to the most commonly used laboratory tests.

The approach of using data from hospitalized patients to produce reference values for laboratory results has been made possible by the introduction of laboratory automation and computerization. From databases of hospital information systems, defined reference subjects can be individually selected by combining large numbers of laboratory, diagnostic, and other classification data. With further development of the techniques needed, acquisition of data from clinical events will become more reliable and comprehensive.

We selected the groups of reference individuals for erythrocyte variables from the total group of hospital-

![Graphs of Hgb, MCV, and RBC distributions for females and males](image-url)

Fig. 2. Distribution of Hgb results from 1786 women and 1454 men "reference patients" (solid line) and those from unselected patients (inpatients, ages 20–65 years; dotted line). Vertical bars mark the 2.5 and 97.5 percentile limits both for unselected (dotted line) and reference patients (solid line) to demonstrate the effect of selection. The reference limits given as numbers are for the selected cases (with the 95% binomial confidence limits).

Fig. 3. Distribution of (left) MCV results and (right) RBC counts from 1784 women and 1450 men "reference patients" (solid lines) and those from unselected patients (dotted lines). Reference limits as shown in Fig. 2. Note differences in scale.
Table 5. Age-related reference intervals (nonparametric estimates with 95% confidence intervals) from hospitalized patients.

<table>
<thead>
<tr>
<th>Age, years</th>
<th>Percentiles (and confidence intervals)</th>
<th>Hemoglobin, g/L</th>
<th>Median</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Women (P = 0.0001)*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–30</td>
<td>109 (104–115)</td>
<td>151 (148–161)</td>
<td>132</td>
<td>274</td>
</tr>
<tr>
<td>30–45</td>
<td>110 (107–114)</td>
<td>151 (150–154)</td>
<td>133</td>
<td>788</td>
</tr>
<tr>
<td>45–55</td>
<td>116 (110–120)</td>
<td>155 (151–159)</td>
<td>135.5</td>
<td>402</td>
</tr>
<tr>
<td>55–65</td>
<td>113 (106–119)</td>
<td>154 (151–168)</td>
<td>136</td>
<td>342</td>
</tr>
<tr>
<td>Men (P = 0.0029)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–30</td>
<td>133 (128–137)</td>
<td>170 (168–181)</td>
<td>152</td>
<td>236</td>
</tr>
<tr>
<td>30–45</td>
<td>129 (122–133)</td>
<td>169 (168–172)</td>
<td>150</td>
<td>477</td>
</tr>
<tr>
<td>45–55</td>
<td>126 (119–130)</td>
<td>169 (167–173)</td>
<td>150</td>
<td>353</td>
</tr>
<tr>
<td>55–65</td>
<td>122 (114–127)</td>
<td>170 (168–178)</td>
<td>149</td>
<td>388</td>
</tr>
</tbody>
</table>

Mean corpuscular volume, fl.

Women (P = 0.0017) |                                      |                 |        |   |
| 20–30      | 82 (78–83)                            | 98 (96–101)     | 90     | 273 |
| 30–45      | 79 (73–81)                            | 99 (98–100)     | 91     | 789 |
| 45–55      | 82 (79–83)                            | 99 (98–101)     | 91     | 402 |
| 55–65      | 81 (74–82)                            | 97 (96–100)     | 90.5   | 340 |

Men (P = 0.0001) |                                      |                 |        |   |
| 20–30      | 84 (82–84)                            | 96 (97–99)      | 90     | 236 |
| 30–45      | 85 (84–86)                            | 101 (100–104)   | 92     | 478 |
| 45–55      | 83 (80–85)                            | 101 (99–104)    | 92     | 350 |
| 55–65      | 83 (82–85)                            | 100 (99–102)    | 91     | 386 |

RBC count (× 10^12/L)

Women (P = 0.0001) |                                      |                 |        |   |
| 20–30      | 3.63 (3.34–3.75)                      | 4.99 (4.83–5.23) | 4.30   | 272 |
| 30–45      | 3.67 (3.58–3.72)                      | 4.95 (4.88–5.04) | 4.31   | 798 |
| 45–55      | 3.68 (3.56–3.79)                      | 5.01 (4.92–5.10) | 4.37   | 402 |
| 55–65      | 3.71 (3.17–3.85)                      | 5.13 (5.00–5.24) | 4.43   | 341 |

Men (P = 0.0001) |                                      |                 |        |   |
| 20–30      | 4.35 (4.02–4.43)                      | 5.58 (5.46–5.82) | 4.87   | 236 |
| 30–45      | 3.93 (3.82–4.11)                      | 5.42 (5.38–5.50) | 4.78   | 478 |
| 45–55      | 3.92 (3.82–4.06)                      | 5.48 (5.40–5.66) | 4.74   | 350 |
| 55–65      | 3.97 (3.71–4.04)                      | 5.51 (5.42–5.70) | 4.74   | 368 |

* Significance of the differences between the four age subgroups (Kruskal-Wallis analysis of variance).

Reliability of Diagnostic Data

Our experimentation with this a posteriori technique revealed several problems. With respect to Hgb results, 18–29% of cases were erroneously included in the reference groups because of errors in the discharge database as compared with patients' records (16). However, these errors were evenly distributed between the low 15%, intermediate 70%, and high 15% percentiles of the total distribution of results (P = 0.096 for women, 0.161 for men, by 2x test). Before applying this selection method, one should investigate the practices by which computerized patient databases are created in different hospitals.

Conceivably, a higher rate of inadequate diagnoses would be found if the patients were reinvestigated (accuracy of clinical diagnoses was not investigated in this study). In another study, Heliövaara et al. (17) found that 22% of the cases with either a stroke or a myocardial infarction were missing from the hospital records in Finland; on the other hand, 9% of codes for myocardial infarction and 19% of those of strokes were not supported by patient interviews or medical records requested by those investigators. In the US and Scotland, comparison of databases with patients' records showed that only ~60–65% of four diagnoses given to patients were reported to have been coded correctly (18, 19).

In the production of reference intervals, a comprehensive list of diagnoses for acute as well as chronic diseases is needed, taking into account even more than four diagnoses. One can use the times of specimen collection to verify the admission status of patients, as opposed to (e.g.) postoperative values. Also, information on medication is essential for the interpretation of many test results in clinical chemistry. In our hospitals, the infor-
formation on drug therapy is not currently being stored in patient databases. Development of electronic patient records including this information should be encouraged.

Refinements of the presented selection criteria (Table 1) may be necessary in the hospitals where the clinical specialties or the variety of diseases differ from those of our hospitals (Table 2). Additional exclusion criteria may be revealed. In the transfer of the criteria, these may be checked by pathogenetic and statistical principles.

Surgical wards often disregarded the repeated recording of diagnosis of a chronic disease. However, despite errors and failures in the recording of diagnoses, reference intervals produced from patients' data may be good approximations of the true intervals, because the differences in the percentiles of misselected cases between the three analyzed Hgb subfractions (low 15, intermediate 70, and high 15 percentiles) were not statistically significant. The greatest rate of misselection was among the very low Hgb results, which may have decreased the estimate for 2.5% limit for women (Fig. 2).

In conclusion, we suggest that the use of computers in storing patient and laboratory data makes it possible to combine individual discharge diagnoses, demographic data, and other relevant information from the treatment periods with laboratory results, thus fulfilling the IFCC recommendation for individually characterized reference subjects. In the future, such an approach may be enhanced by structured electronic patients' records, which will allow more detailed classifications of patients. The distributions of reference values produced may then be utilized in a new way to compare patients' results with those of "healthy" and "diseased" reference populations of the same hospital. When clinical doctors discover the benefits of accurate databases created by their own efforts, they will be motivated to improve the techniques and practices in coding and storing patients' data.

We acknowledge the support of this study by Turku University Hospital and NORDKEM, Helsinki.

References