Interlaboratory variability of bilirubin measurements


During an 8-month study, 14 laboratories used automated analytical systems to measure total bilirubin concentrations in lyophilized bovine specimens containing 38, 169, and 253 μmol/L bilirubin (2.2, 9.9, and 14.8 mg/dL, respectively). The measured mean ± SD (n, range) were: 39 ± 7 μmol/L (n = 90, 31–53) [2.3 ± 0.4 mg/dL (1.8–3.1)]; 176 ± 29 μmol/L (n = 89, 146–222) [10.3 ± 1.7 mg/dL (8.5–13.0)]; and 260 ± 43 μmol/L (n = 103, 208–316) [15.2 ± 2.5 mg/dL (12.1–18.5)]. In comparison with target values, measurements were consistently lower at 4, higher at 6, and within ±4% at 4 laboratories for each of the three concentrations. The measured values for each concentration remained fairly constant during the study at each laboratory. We conclude that bilirubin measurements differed significantly from the established target values at most of the participating laboratories.

INDEXING TERMS: control material • quality control • automated analysis

One of the laboratory tests most commonly performed on newborn infants is the serum total bilirubin measurement, because jaundice occurs in most neonates during the first several days after birth. This clinical syndrome results from the combined effects of: (a) a natural rate of bilirubin production that is greater per body weight basis than for adults; (b) an abrupt cessation of bilirubin clearance by the placenta at the time of birth; (c) a transient impairment of bilirubin conjugation by the neonatal liver; and (d) enterohepatic circulation of bilirubin, which decreases gradually after birth [1]. Pediatric care practitioners depend on determinations of total bilirubin (TBil) to make decisions about phototherapy or exchange transfusions.15 Furthermore, the recently published guidelines from the Provisional Committee for Quality Improvement and Subcommittee on Hyperbilirubinemia [2] rely on an algorithm supplemented by a table of TBil concentrations that indicate when phototherapy and (or) exchange transfusion should be initiated. Thus, the accurate determination of serum or plasma TBil concentrations is important for the management of hyperbilirubinemia in newborns.

Apparent interlaboratory variability in bilirubin quantification can sometimes reflect differences in what is being reported, such as TBil vs "neonatal" TBil, rather than differences in...
laboratory accuracy [3]. However, even when the same determination is being made, the results reported by clinical laboratories may vary [4, 5]. In the course of determining bilirubin production in healthy term infants through breath carbon monoxide measurements, 14 different clinical laboratories also measured TBil in the same subjects [6]. To assess interlaboratory variability, each of the laboratories also made periodic measurements of bilirubin in bovine serum albumin-based material containing three concentrations of bilirubin. Here we report their findings.

**Materials and Methods**

Each clinical laboratory was supplied with lyophilized buffered bovine albumin (Bilirubin Standard Set; Sigma Chemical Co., St. Louis, MO) containing bilirubin concentrations of 38, 169, and 253 μmol/L (2.2, 9.9, and 14.8 mg/dL, respectively). The bilirubin concentrations had been determined spectrophotometrically by the manufacturer on the basis of the molar absorptivity of alkaline azobilirubin (procedure no. 605). The SD for this determination was 5%. At the beginning of the study, material from one production lot (no. 31H-6145, exp. date April 1993) with a 2-year stability was supplied in a randomized, blinded way to each laboratory, where it was stored in the dark at 2–6 °C. At each laboratory, the material was reconstituted with 3.0 mL of distilled water immediately before use on the first Wednesday of each month, except for the first month of the study, when specimens were prepared for weekly analysis. Local laboratory procedures and study results were not adjusted on the basis of control sample results. Five to nine blinded measurements of each concentration were made over a period of 8 months (November 1991 through August 1992) with the analytical system in use at each laboratory.

Nine institutions determined bilirubin with the Kodak EktaChem system (Johnson and Johnson Clinical Products, New Brunswick, NJ), an automated clinical chemistry analyzer. This system allows for measurement of serum bilirubin by either of two dry slide methods: TBil, based on a modified diazo method, and neonatal bilirubin (NBil), a dual-wavelength colorimetric method that directly measures the unconjugated (Bu) and conjugated (Bc) bilirubins (Bu + Bc = NBil) [3]. For the purpose of this study, only TBil measurements were made and reported because not all laboratories were able to measure NBil. Although a previous study reported a statistically significant difference (P < 0.0001) between TBil and NBil, this difference was of no clinical significance (163 ± 70 vs 165 ± 69 μmol/L) [3].

Hitachi instruments (Models 717, 736, and 737; Boehringer Mannheim Corp., Laboratory Systems, Indianapolis, IN), suitable for measuring TBil in neonate and adult plasma or serum, were used in three laboratories. These instruments use the diazo reagent 2,5-dichlorophenyl diaziridinio tetrafluoroborate, which reacts rapidly with bilirubin under acidic conditions. A detergent is included to accelerate the reaction and to avoid protein precipitation [7, 8]. The light absorption of the azobilirubin is measured during the reaction time with a fixed diode-array spectrophotometer at primary and secondary wavelengths of 570 and 600 nm, respectively. These instruments were calibrated with saline and a multiconstituent material, Precital™ calibrator serum (Boehringer Mannheim Corp.).

Two laboratories used the Paramax 720 Analytical System (Baxter Diagnostics, Scientific Products Div., McGaw Park, IL). This TBil technique is based on the solubilization of bilirubin by a surfactant [9, 10], followed by diazotization of sulfanilic acid to react with bilirubin to form azobilirubin chromogen. The absorbance of the reaction product is measured bichromatically at 550 and 630 nm. The reagents necessary for the endpoint determination of TBil are contained within a single tablet. This system was calibrated at least every 90 days, or sooner if there were changes in reagent lots or controls were outside acceptable limits, with Paramax® Bilirubin calibrators I and II.

Data are presented as means ± SD and CVs. One-way analysis of variance was used to determine interlaboratory variability.

**Results**

The mean ± SD, grouped by analytical system used, of the three materials measured at each laboratory are given in Table 1. Fig. 1 summarizes the results for each of the materials at each of the laboratories. Six of the 14 laboratories reported values that were consistently >104% of target values for all three samples; 4 laboratories showed consistently lower values (<96% of target values).

The within-laboratory variability across time, expressed as CV, ranged from 1.3% to 15.8% (for 9 laboratories, <5.0%), from 1.4% to 15.4% (for 10 laboratories, <5.0%), and from 2.0% to 17.2% (for 7 laboratories, <5.0%) for the 38, 169, and 253 μmol/L materials, respectively.

For 6 of the 90 measurements (6.7%) of the 38 μmol/L (2.2 mg/dL) material, the results exceeded 51 μmol/L (3.0 mg/dL); all 6 occurred at one site. For the 89 measurements of the 169 μmol/L (9.9 mg/dL) material, 25 (28.1%) exceeded 205 μmol/L (12.0 mg/dL), and 36 of the 103 measurements (35.0%) of the 253 μmol/L (14.8 mg/dL) sample exceeded 274 μmol/L (16.0 mg/dL). One-way analysis of variance showed significant differences in results for all three concentrations between the 14 laboratories (P < 0.01).

The mean measured TBil values for each concentration measured at all laboratories remained fairly constant over the 8 months of study (Fig. 2). The absence of any time-related trend indicated a generally consistent internal quality control at the laboratories.

**Discussion**

Like Schreiner and Glick [4], who also studied interlaboratory bilirubin measurement variability, we conclude that little or no decrease in interlaboratory bilirubin variability has occurred since 1982, the date they made the same comment in reference to the 10 years preceding their article. Some of the variation in results may be attributable to the origin of the specimen [11]. The Kodak EktaChem manual recommends that specimens from neonates <14 days old should not be analyzed for TBil because differences of ±10% will be observed across the concentration range. However, the results of a study by Langbaum et al. [3] did not support this recommendation. The reason for this disagree-
Table 1. TBil concentrations (µmol/L) in lyophilized material (Levels I, II, and III) measured at 14 different clinical laboratories with three types of analytical systems over 8 months.

<table>
<thead>
<tr>
<th>Lab. no.</th>
<th>Level I (38 µmol/L)</th>
<th>Level II (189 µmol/L)</th>
<th>Level III (253 µmol/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Kodak Ektachem</td>
<td>2</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>33</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>37</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>37</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>39</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>44</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>45</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>45</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>53</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>7</td>
<td>58</td>
</tr>
<tr>
<td>Hitachi</td>
<td>8</td>
<td>39</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>46</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>44</td>
<td>4</td>
<td>19</td>
</tr>
<tr>
<td>Paramax</td>
<td>1</td>
<td>31</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>31</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>All labs.</td>
<td>39</td>
<td>7</td>
<td>90</td>
</tr>
</tbody>
</table>

ment is apparently not known. We have no information available indicating that our samples would be subject to similar interference.

The deviation from the target bilirubin concentration did not appear to depend on the measurement technique because the instrument most commonly used (Kodak Ektachem) measured low (n = 2), on-target (n = 3), and high (n = 4) to nearly the same extent (see Table 1 and Fig. 1). The SD of the control sample bilirubin measurements was large between laboratories and was sustained over time, as shown in Fig. 2.

Unfortunately, the interlaboratory variability may have consequences for clinical practice on a case-by-case basis. Although TBil measurements were nearly always consistent at each laboratory for each of the three concentrations (Fig. 1), at most of the laboratories the results differed by >4% from the control materials' nominal target values established by Sigma. This variability was sustained over time and is important. Only one laboratory (no. 10; not shown) reported an abrupt and sustained increase in the measured bilirubin concentrations at each of the three concentrations, which brought the values closer to the target values. No records are available to determine if a change of reagent was done, which might explain the increase. Laboratory 9 also reported a wide variation of concentrations for each material, but the measurements appeared to vary at random over time (not shown). Thus, because the within-laboratory variability is relatively low, the real problem exists between laboratories, even between those using the same brand of instrument—perhaps because of inconsistent calibration locally.

The disturbing fact remains that, because clinical guidelines depend heavily on bilirubin action values, the under- and overestimation of TBil concentrations could lead to withholding of therapy or unnecessary clinical intervention in some newborns. For most of the healthy newborn population, this would result in unnecessary blood sampling and phototherapy, which would represent a nuisance and unnecessary cost, rather than increased risk [12, 13]. However, an increased likelihood that the variability described could result in unnecessary exchange transfusions in infants with higher TBil concentrations would involve a risk more dangerous than simply one of nuisance or unnecessary cost. This is underscored by the aforementioned practice parameters, which were developed in "the general belief that therapeutic interventions for hyperbilirubinemia in the healthy term infant may carry significant risk relative to the uncertain risk of hyperbilirubinemia in this population" [2].

The debate continues about the relationship of serum TBil concentrations and the risk of kernicterus under various conditions [14, 15]. Clearly, not all infants whose results are above any selected action concentration value will develop permanent injury in association with hyperbilirubinemia, nor are infants with bilirubin concentrations below a certain threshold immune from injury associated with hyperbilirubinemia [14]. The cause of hyperbilirubinemia is probably multifactorial and contributes to the risk of kernicterus, especially in very-low-birth-weight infants [1]. Also, study of the duration of exposure to hyperbilirubinemia and the risk of kernicterus has been inadequate. Further, sudden decreases in serum TBil concentration may not signify an improving condition, but rather a worsening one, with an attendant greater risk for permanent injury as the bilirubin distributes into tissues [16]. Another possibility, however, is that
some of the lack of positive correlation may be caused by interlaboratory variability in bilirubin measurements.

Although matrix effects might explain some of the apparent method-specific biases we noted [17, 18], the widely disparate results from the Ektachem instruments, for which any matrix-related bias would be expected to be more or less constant, suggest very serious calibration problems with these instruments in the field. Thus, current neonatal bilirubin measurements may present a real problem in laboratories across the country. However, several other errors are possible, including the accuracy of aliquoting the study material before lyophilization, decay of bilirubin during inappropriate storage, and inaccurate reconstitution procedures.
We conclude that, despite the recommendations made by Schreiner and Glick, no systematic distribution of bilirubin solutions for quality control and calibration in the range appropriate for the neonatal population is used routinely in clinical laboratory practice. Many accredited laboratories participate in a quality survey program administered by the College of American Pathologists (CAP) [19], and the yearly results are published as CAP Surveys. For 1994, analysis of total bilirubin samples by 567 clinical laboratories using a plethora of techniques (including those used for this study) yielded CVs between 5.6% and 11.0% for survey samples containing 21–81 μmol/L (1.25–4.76 mg/dL) bilirubin. However, proficiency testing programs that use artificial materials prone to matrix-induced biases, such as the CAP Surveys, do not unequivocally assess laboratory accuracy. In addition, we recommend such programs should frequently offer proficiency testing challenges at high bilirubin concentrations, which are in the range of clinical interest for neonatal bilirubin. Until widely available standardization processes free of matrix effects are introduced, accurate and precise bilirubin measurements in clinical laboratories cannot be assured; this fact must be taken into consideration, especially when pediatricians are mandated to respond to proposed bilirubin action guidelines. Finally, the need for accurate bilirubin determinations assumes greater importance and meaning in view of current early hospital discharge policies.

This study was supported by a grant from the National Institutes of Health, National Institute of Child Health and Human Development, through cooperative agreements with the institutions identified by superscripts (see authors’ affiliations): 1 U10 HD27880, 2 U01 HD19897, 3 U10 HD27904, 4 U10 HD21364, 6 U10 HD27856, 7 U10 HD21385, 8 U10 HD21373, 9 U10 HD21415, 10 U10 HD21373, 11 U10 HD27851, 12 U10 HD27881, 13 U10 HD27853, and 14 U10 HD27871.

References


