

## Errors in Measurement of Total Bilirubin: A Perennial Problem

In 1960, Mather [1] stated that "bilirubin determinations are perhaps the most notoriously unreliable of any in clinical chemistry." Twenty-two years later, Watkinson et al. [2] came to much the same conclusion. Now, the study by Vreman et al. [3] reveals that accuracy in measuring serum total bilirubin can be elusive even in university hospitals. It is, indeed, disturbing to see bilirubin results seemingly so unreliable that for one specimen the reported concentrations ranged from 110 to 210 mg/L (188 to 359  $\mu\text{mol/L}$ ) when all of the laboratories' results are included. In this editorial, we will attempt to identify potential sources of real and apparent inaccuracy and imprecision in bilirubin analyses and try to explain some of the findings reported by Vreman et al.

To begin with, bilirubin values of the control specimens used in their study are assigned, according to Sigma Chemical Co. (St. Louis, MO), by the Jendrassik-Gróf method and use of a molar absorptivity ( $\epsilon$ ) value of 73 000  $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$  for the alkaline azopigment at 600 nm [4]. Subsequent measurements of the  $\epsilon$  of the alkaline azopigment of the National Institute of Standards and Technology (NIST) bilirubin, Standard Reference Material (SRM) 916, gave a mean  $\epsilon$  value of 75 500  $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$  at 598 nm [5]. Thus, bilirubin values assigned by Sigma could be overestimated by 3.3%. The corrected values for the three control specimens should be 36.7  $\mu\text{mol/L}$  (21.5 mg/L), 163.4  $\mu\text{mol/L}$  (95.6 mg/L), and 244.6  $\mu\text{mol/L}$  (143 mg/L) for levels I, II, and III, respectively.

Bilirubin SRM 916 from NIST is the Reference Material for calibrating methods for bilirubin; secondary calibrators should be traceable to the SRM 916. The SRM 916, being unconjugated bilirubin, is insoluble in water near physiologic pH, but soluble at alkaline pH. Aqueous solutions of bilirubin are stabilized by addition of protein, e.g., human serum, bovine serum, human serum albumin (HSA), or bovine serum albumin (BSA). The accuracy of bilirubin standard solutions is verified by analysis with the Reference Method and calculation of the  $\epsilon$  of the alkaline azopigment at 598 nm. The established  $\epsilon$  value is 75 500  $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$  with an SD of 550 [5]. Because the SD is very small, we propose that a range of 75 500  $\pm$  1900 ( $\pm$ 2.5% of the mean value) be used as a criterion of accuracy of bilirubin calibrators. Additional criteria for accuracy should be the  $\epsilon$  values of bilirubin in caffeine reagent at 432 and 457 nm; the reported  $\epsilon_{432\text{nm}}$  and  $\epsilon_{457\text{nm}}$  mean values and ranges are 49 500 ( $\pm$ 1300) and 48 800 ( $\pm$  1300)  $\text{L}\cdot\text{mol}^{-1}\cdot\text{cm}^{-1}$ , respectively [6].

With the Reference Method, which is based on the Jendrassik-Gróf principle, the  $\epsilon$  value at 598 nm is the same in the most common protein matrices (human serum, HSA, or BSA). This is also the case for the  $\epsilon$  values of bilirubin at 432 and 457 nm in the same protein matrices. Therefore, bilirubin calibrators may be prepared in any of these protein matrices [5]. This may not hold true for other bilirubin methods. If matrix effects are suspected or documented, values for bilirubin in calibrators made in HSA or BSA must be assigned by comparison to bilirubin calibrators in human serum. It is strongly recommended that the effect of protein matrix be evaluated before bilirubin methods are placed in use.

The report by Vreman et al. shows some unexpected and disturbing findings. Their Fig. 1 shows that within-laboratory variability is extremely large in some of the laboratories (e.g., #9,

10, 13, and 14), but relatively small in others (e.g., #1, 3, 5, 7, and 8). Small within-laboratory variability in at least some of the laboratories suggests that a large vial-to-vial variability of bilirubin in the Sigma controls is an unlikely source of the problem. It is hard for us to understand the extremely large variability with control level III in laboratories #9, 10, 13, 14, and with level II in laboratories #9 and 10, without suspecting lapses in internal quality control, failure to calibrate instruments properly with reagent lot changes, or perhaps improper reconstitution or handling of vials.

Let us now examine the accuracy of the various methods. Bilirubin values from laboratories #1 to 4 were consistently low for all three levels by  $\sim$ 10–15%. The most likely causes for what appears to be a proportional negative bias are: (a) some sort of incorrect instrument calibration (e.g., bilirubin concentrations of the individual instrument's calibrator fluids were actually higher than those assigned by the instrument manufacturer), (b) the bilirubin in the Sigma controls, made with BSA, was somehow less "reactive" than bilirubin in human serum or the bilirubin in the instrument's calibrator, or (c) the  $\epsilon$  value of the azopigment could be lower in BSA than in human serum. From the data shown by Vreman et al., it is difficult to differentiate among these three possible causes.

The data from the Ektachem, the most commonly used instrument in the study, are particularly puzzling to us. Two of the Ektachems (#2 and 4) gave consistently low results, three (#5, 6, and 7) were reasonably near Sigma's assigned values, and four (#9, 10, 11, 14) were consistently high. In addition, several of the Ektachem's results (#10, 13, 14) seemed unexpectedly variable within a given laboratory through the 8 months of the study. Again, inadequate attention to instrument calibration and failure to check instrument performance across reagent (slide) lots are possible explanations. Long-term (1 year, including several slide lot changes and recalibrations) CVs from Ektachem analyzers in our own laboratories were  $\sim$ 8–10% at a bilirubin concentration of 11 mg/L, and 2–3% at 180 mg/L. We believe that our intralaboratory CV of 2–3% is more consistent with the interlaboratory CVs in the College of American Pathologists (CAP) surveys (see below) than with those obtained by some of the Ektachem analyzers in the Vreman study.

Table 1 shows proficiency testing data from two of the CAP surveys. CAP specimens NB02, NB10, and NB15 came from a single HSA-based, lyophilized control pool. They were distributed in April, August, and November 1995, respectively, to the  $\sim$ 800 laboratories participating in the CAP Neonatal Bilirubin survey. Specimens NB05, NB11, and C96 were from a different but similarly prepared pool. They were sent in April 1995 as NB05 and NB11 to the laboratories participating in the Neonatal Bilirubin survey and in November 1995 as C96 to the  $\sim$ 6000 laboratories participating in the CAP Chemistry survey. Three of the instruments listed in Table 1 (DuPont *aca*, Ektachem, and Paramax) can analyze bilirubin by either direct spectrophotometry or diazo methods; the other three (DuPont Dimension, BMC/Hitachi, and Beckman Synchron), to our knowledge, use only diazo methods.

The within-method variability in the CAP data is quite low. Except for specimen C96, the CVs range from 1.4% to 4.4%,

**Table 1. Bilirubin mean values (mg/L) and CVs (%) for selected instruments reported in the CAP 1995 Neonatal Bilirubin (NB) and Chemistry (C) surveys.**

Method*	NB02		NB10		NB15		NB05		NB11		C96	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
All laboratories	190.5	8.2	189.4	8.1	189.9	8.1	147.9	8.3	145.0	7.9	150.2	12.9
DuPont <i>aca</i> (DS)	191.1	2.4	192.6	3.3	188.5	2.4	146.7	3.1	145.1	3.1	155.0	2.1
DuPont Dimension (DZ)	224.7	2.3	223.8	2.4	224.5	1.4	172.6	1.9	169.3	1.7	173.4	2.4
Beckman Synchron (DZ)	191.9	2.6	193.0	3.5	193.1	3.1	148.6	2.0	145.2	2.0	148.0	3.0
BMD/Hitachi (DZ)	212.6	3.3	213.9	3.7	211.8	2.6	164.2	3.5	160.5	3.4	165.3	3.6
Ektachem/NBIL (DS)	193.5	3.8	189.3	2.7	190.3	3.6	153.0	3.0	148.4	3.2	147.3	5.1
Ektachem/TBIL (DZ)	192.8	4.0	186.8	2.7	190.9	2.6	152.6	3.6	147.0	2.9	125.0	8.8
Dade Paramax (DS)	183.6	4.4	182.1	3.0	181.6	2.9	141.5	2.5	137.0	2.7	155.2	3.8

\* DS, direct spectrophotometry; DZ, diazo methods.

Specimens NB02, NB10, and NB15 are from a single pool. Specimens NB05, NB11, and C96 are also from a single (different) pool.

—and these are interlaboratory CVs. Because individual laboratories in CAP surveys code their own methods, it is often difficult to be sure that the method codes are correct. For example, some of the laboratories that analyzed C96 by use of the Ektachem TBIL method may have incorrectly coded it as the Ektachem spectrophotometric NBIL ( $B_u/B_c$ ) method and vice versa. We suspect that few of the laboratories in the Neonatal Bilirubin survey actually used the Ektachem TBIL slide method, because the manufacturer specifically recommends against its use with neonatal specimens. On the other hand, laboratories in the Chemistry survey could have used either the Ektachem TBIL method or the NBIL method. With two possible Ektachem methods, miscoding the Ektachem NBIL as Ektachem TBIL and vice versa becomes likely. This miscoding, combined with the method-specific means for the Ektachem's TBIL and NBIL methods, would explain the unusually high CVs for specimen C96 in the Chemistry survey, compared with the CVs for the same specimens in the Neonatal Bilirubin survey.

The method-specific means in the CAP surveys provide, assuming there are no matrix effects, some sense of each method's accuracy. Assuming that the "all-laboratory" grand mean values (190 and 148 mg/L for specimens NB02/NB10/NB15 and NB05/NB11/C96, respectively) are close to the true values of bilirubin, the Hitachi shows in the CAP survey an average bias of +11%, which is much less than the bias (20–29%) shown by two of the three Hitachi instruments in the Vreman study; the third instrument shows a negligible bias ranging from 1.4% to 1.8%. With one instrument being accurate and two being inaccurate, we suspect problems with calibration, especially in view of the Hitachi values in the CAP surveys. The Paramax shows a positive bias of 5% with specimen C96, which contrasts the negative biases (8.5–11.5%) seen in the Vreman study.

The Ektachem methods show virtually no bias in the NB series of specimens distributed in the CAP Neonatal Bilirubin survey. However, results for specimen C96 in the CAP Chemistry survey are more difficult to explain. Laboratories that use

the Ektachem NBIL ( $B_u/B_c$ ) method for bilirubin are supposed to report results using the Ektachem/NBIL method code, whereas those using the Ektachem diazo-based TBIL slide are to report under the Ektachem/TBIL code. What is rather peculiar is that there is very little difference between these two Ektachem methods for specimens NB05 and NB11, but for C96, which was sent out in the CAP Chemistry survey only 1 week after NB11, there was a significant negative bias for the Ektachem/TBIL method (mean, 125 mg/L or 214  $\mu\text{mol/L}$ ), compared with the Ektachem/NBIL (mean, 147 mg/L or 251  $\mu\text{mol/L}$ ). The most likely explanation for this apparent anomaly is that there was significant method miscoding by those in the Kodak/TBIL method group of the Neonatal Bilirubin survey. The Ektachem manual specifically recommends that the TBIL slide should not be used on neonatal specimens. Thus, we hypothesize that most of the Neonatal Bilirubin survey participants used the NBIL method, regardless of the method code under which they reported. In the Chemistry survey, in which C96 was circulated, both methods were likely used. The fact that the method-specific CVs for the two Ektachem methods on C96 were much higher than others suggests to us that there was a fair amount of miscoding between the two Ektachem methods for C96 too.

One of the authors (B.T.D.) and a colleague (Roberta G. Reed, Cooperstown, NY) analyzed in their laboratories C96 with the Ektachem TBIL and NBIL methods and by the Reference Method. Results for NBIL (147 and 149 mg/L) and TBIL (129 and 110 mg/L) were close to those obtained for C96 by the Ektachem in the CAP survey. The bilirubin concentration by the Reference Method was 143 mg/L. Analysis of bilirubin standards prepared in BSA (Cohn Fraction V), HSA (purified by affinity chromatography), HSA Cohn Fraction V, and human serum led to the following conclusions: (a) with bilirubin in human serum, the Ektachem TBIL value was equal to the value of the Reference Method and 95% of the NBIL value; (b) with bilirubin in all of the other proteins, TBIL values were equal to the Reference Method values, but from 15% to 20% lower than the NBIL values.

These findings strongly suggest a matrix effect with the Ektachem TBIL and NBIL methods when matrices other than human serum are used, but do not explain why the TBIL value for C96 is lower than the value by the Reference Method. The term "matrix effects" is used to denote a special type of interference that is not ascribable to a specific chemical substance. Matrix effects are encountered in biological specimens in which certain physical or chemical properties have been altered during their preparation; e.g., addition of stabilizers and preservatives, dialysis, and freeze-drying, to name but a few. In some analytical systems these materials, used for calibration or for monitoring precision and accuracy, do not behave like fresh patients' samples. That is, an analytical system that provides accurate results in patients' samples may fail to provide accurate results in processed biological materials [7,8].

Some methods for measuring bilirubin are very sensitive to matrix effects. For example, with diazo methods in which methanol is used as the accelerator, the "reactivity" of bilirubin varies with the protein matrix; it is highest in human serum, followed by BSA and HSA [9]. The same is true for the reactivity of ditaurobilirubin (the bilirubin conjugate often used for calibrating methods for direct bilirubin) in some diazo methods for direct bilirubin [10]. Direct spectrophotometric methods are also affected by the protein matrix. Calibrators made in matrices other than human serum may not be suitable for calibrating direct spectrophotometric methods because the absorption maximum and the  $\epsilon$  value of unconjugated bilirubin vary with the protein matrix and with the process used to prepare the protein. For example,  $\epsilon$  values of unconjugated bilirubin in crystalline HSA or BSA are much higher than in the corresponding Cohn Fractions V or in human serum [11, 12]. Surface-active agents (e.g., Brij®35) and the albumin fatty acid content increase the absorptivity of unconjugated bilirubin and shift its absorption maximum to longer wavelengths [11, 13]. These effects on the absorption spectra and  $\epsilon$  of unconjugated bilirubin are completely abolished by substituting caffeine reagent for phosphate buffer in the direct spectrophotometry of bilirubin [6, 14]. The situation could get more complicated by the presence of ditaurobilirubin in calibrators and control sera, the spectrum and  $\epsilon$  value of which are also matrix-dependent [10].

For these reasons, therefore, a given control or calibrator preparation may appear to yield more than one "true" value for bilirubin, such that method-dependent assigned bilirubin concentrations are necessary for artificially prepared materials that are to be analyzed in clinical laboratory instruments. As an aside, we believe that matrix effects for total bilirubin could be nearly eliminated, at least for methods involving liquid reagents, if all methods were to use caffeine-benzoate-acetate as the promoter. Although the question of whether matrix effects are responsible for some of the observed variability of Ektachem results reported by Vreman et al. cannot be answered with certainty, the lack of consistency in the direction and magnitude of the bias tend to rule against matrix effects. The numbers of the other two analyzers that were used in Vreman's study are too small to draw any meaningful conclusions.

The term "neonatal bilirubin" usually denotes bilirubin measured in the blood of neonates by direct spectrophotometry. Such methods were developed 30 to 40 years ago to avoid falsely low bilirubin values by diazo methods with methanol as the accelerator, which were very popular at that time in the US, but very susceptible to hemoglobin interference. The measurement

is based on the assumption, which is true in most neonates, even those with physiological or pathological hyperbilirubinemia, that unconjugated bilirubin is the predominant bile pigment in blood. Neonatal bilirubin is thus really not a separate analyte; it is identical to the unconjugated bilirubin found in adult sera. Since suppression of total bilirubin values by hemolysis can be minimized by performing the coupling reaction near pH 7 (e.g., the Jendrassik-Gróf principle, but omitting the addition of alkaline tartrate), values by such diazo methods should match those of direct spectrophotometry if both kinds of methods are properly calibrated.

Preparing bilirubin calibrators in human serum used to be common practice in clinical laboratories two to three decades ago. With the advent of automation, commercial vendors have become the major source of reagents and most calibrators. Thus, the primary responsibility for accuracy has shifted to instrument and reagent manufacturers. We believe that the best calibrator for total bilirubin methods is unconjugated bilirubin in human serum, the bilirubin concentration of which has been accurately assigned by analysis with the Reference Method [5]. For most of the methods based on the Jendrassik-Gróf principle, calibrators made in HSA or BSA are equivalent to those made in human serum. For other methods, calibrators made in nonhuman serum matrices should have bilirubin values assigned by comparison with calibrators in human serum. Monitoring of the analytical quality of bilirubin measurements is far more important when analyzing specimens from neonates, because of the risk of kernicterus and irreversible brain damage, than specimens from adults in whom bilirubin per se is harmless. To ensure that laboratory results are reliable would require daily use of quality-control materials with high (~200 mg/L, 340  $\mu$ mol/L) bilirubin concentrations, but also periodic use of proficiency testing challenges with similarly high (e.g., CAP Neonatal Bilirubin survey specimens) and preferably accurately known bilirubin concentrations.

Freeze-drying of serum causes denaturation of lipoproteins, usually giving substantial turbidity upon rehydration, which without proper blanking introduces a serious positive bias [15]. Turbidity can be prevented by substituting HSA or the less expensive BSA for human serum. Such optically clear preparations with total bilirubin values assigned by the Reference Method would be suitable for establishing whether inaccuracies observed with certain diazo methods are the result of matrix effects or inadequate calibration. Finally, we believe that having the total bilirubin results of bilirubin proficiency testing programs graded against a single, accurate bilirubin target value traceable directly to the Reference Method value, with exceptions granted only to methods that have proven matrix effects in quantitative terms, would be a valuable adjunct to improving the quality of bilirubin assays in clinical laboratories. Perhaps then the perennial goal of accurate bilirubin results in general clinical practice may finally come into our grasp.

#### References

1. Mather A. Reliability of bilirubin determinations in icterus of the newborn infant. *Pediatrics* 1960;26:350-4.
2. Watkinson LR, St. John A, Penberthy LA. Investigation into paediatric bilirubin analyses in Australia and New Zealand. *J Clin Pathol* 1982;35: 52-8.
3. Vreman HJ, Verter J, Oh W, Fanaroff AA, Wright LL, Lemons JA, et al. Interlaboratory variability of bilirubin measurements. *Clin Chem* 1996;42: 869-73.
4. Gambino RS, Di Re J. Bilirubin assay. Chicago, IL: American Society of Clinical Pathologists, 1968.

5. Doumas BT, Poon PKC, Perry BW, Jendrzczak B, McComb RB, Schaffer R, et al. Candidate reference method for determination of total bilirubin in serum: development and validation. *Clin Chem* 1985;31:1779-89.
6. Vink KLJ, Schuurman W, van Gansewinkel R. Use of caffeine reagent in direct spectrophotometry of bilirubin. *Clin Chem* 1986;32:1389-93.
7. Miller WG, Kaufmann HW. College of American Pathologists Conference XXIII on Matrix Effects and Accuracy Assessment in Clinical Chemistry: introduction. *Arch Pathol Lab Med* 1993;117:343-4.
8. Powers DM. Overcoming the limitations of artificial clinical specimens: linkage to performance of patients' specimens [Editorial]. *Clin Chem* 1995;41:1689-91.
9. Doumas BT, Perry BW, Sasse EA, Straumfjord JV Jr. Standardization in bilirubin assays: evaluation of selected methods and stability of bilirubin solutions. *Clin Chem* 1973;19:984-93.
10. Doumas BT, Wu TW, Poon PKC, Jendrzczak B. Chemical nature of a synthetic bilirubin conjugate and its reactivities in the total and direct reactions by the Jendrassik-Gróf methods. *Clin Chem* 1985;31:1677-82.
11. Doumas BT, Perry BW, Sasse EA. The effect of protein base on the molar absorptivity and absorption spectra of bilirubin [Abstract]. *Clin Chem* 1973;19:647.
12. Wennberg RP, Cowger ML. Spectral characteristics of bilirubin bovine albumin complexes. *Clin Chim Acta* 1973;43:55-64.
13. Reed RG. Kinetics of bilirubin-albumin binding. *J Biol Chem* 1977;252:7483-7.
14. Vink KLJ, van Dreumel RJ, Schuurman W, Wikkeling H, Gransewinkel R, Phielix CJ, et al. A candidate standard for use in calibration of total bilirubin in serum. *Clin Chem* 1987;33:1817-21.
15. Lott JA, Doumas BT. The "direct" and total bilirubin tests: contemporary problems. *Clin Chem* 1993;39:641-7.

**Basil T. Doumas**

*Medical College of Wisconsin*

*P.O. Box 26509*

*Milwaukee, WI 53226*

*Fax 414-257-7899*

*E-mail bdoumas@post.its.mcw.edu*

**John H. Eckfeldt**

*University of Minnesota Hospital and Clinics*

*Laboratory Medicine and Pathology, UMHC Box 609*

*420 Delaware St. SE*

*Minneapolis, MN 55455-0392*

*Fax 612-625-6994*

*E-mail eckfe001@maroon.tc.umn.edu*