

Performance of Bilirubin Determinations in US Laboratories—Revisited

STANLEY F. LO^{1,2*} BASIL T. DOUMAS,^{1,2} and EDWARD R. ASHWOOD³

Background: The diagnosis and management of hyperbilirubinemia in the newborn requires accurate and precise bilirubin determinations. We evaluated the current status of bilirubin measurements in US laboratories by examining data submitted by laboratories participating in the College of American Pathologists (CAP) Neonatal Bilirubin (NB) and Chemistry (C) Surveys.

Methods: We analyzed specimens from the CAP NB and C Surveys by the reference method for total bilirubin in three laboratories. The reference method bilirubin values were compared with bilirubin values reported by Survey participants.

Results: The imprecision (CV) for all instruments combined (CAP-All Instruments) ranged from 4.7% to 5.6% at the bilirubin concentrations tested. The CVs of the four most commonly used instruments were smaller, ranging from 1.9% to 4.5%. Differences between bilirubin values by the reference method and mean values from the four most common instruments ranged from –21.6% to 10.9%. When the same specimens from the NB Surveys were used in the C Surveys, the Vitros values were strikingly different from those of the NB Surveys. The use of different methods in the NB and C Surveys coupled with the presence of a nonhuman protein base and ditaur bilirubin (DTB) in the Survey specimens accounted for the discrepant results.

Conclusions: The evaluation of accuracy is impossible from the CAP Surveys because the specimens consist of bovine serum containing a mixture of unconjugated bilirubin and DTB. For the evaluation of accuracy, we

recommend that Survey specimens consist of human serum enriched with unconjugated bilirubin.

© 2004 American Association for Clinical Chemistry

Hyperbilirubinemia in healthy newborn infants is a common condition treated by pediatric practitioners. Guidelines for the management of hyperbilirubinemia in healthy term newborns were published in 1994 by the American Academy of Pediatrics (1). The guidelines rely heavily on the ability of the physician to recognize jaundice and on the measurement of serum total bilirubin. Thus, the importance of precise and accurate bilirubin results is critical to proper management of hyperbilirubinemia.

Historically, interlaboratory agreement of bilirubin measurements has not been good (2). In 1995, Vreman et al. (3) conducted a study on the status of bilirubin measurements in 14 university hospital laboratories in the US. They used commercial lyophilized controls consisting of bovine serum albumin (BSA)⁴ enriched with unconjugated bilirubin (UBIL). The instruments used in the study were the Kodak Vitros (then Ektachem; n = 9 sites), Hitachi (n = 3), and Paramax (n = 2). The mean values from the Vitros and the Hitachi instruments were, respectively, 5% and 11% higher than the values assigned by the manufacturer of the lyophilized controls; those from the Paramax were lower by 15%. In the College of American Pathologists (CAP) Survey of the same year, the CV for all Vitros participants was 3.3%, whereas in the study by Vreman et al. (3), the Vitros data had a CV of 17%. The high imprecision reported by Vreman et al. was surprising because according to the authors all nine Vitros instruments used the total bilirubin (TBIL) method (TBIL slide) for measuring bilirubin. Similarly the imprecision (CV ~11%) of the three Hitachi analyzers was approxi-

¹ Reference Standards Laboratory, Department of Pathology and Laboratory Medicine, Children's Hospital of Wisconsin, Milwaukee, WI.

² Department of Pathology, Medical College of Wisconsin, Milwaukee, WI.

³ ARUP Laboratories, Department of Pathology, University of Utah, Salt Lake City, UT.

*Address correspondence to this author at: Reference Standards Laboratory, Children's Hospital of Wisconsin, 9000 W. Wisconsin Ave., Milwaukee, WI 53226. Fax 414-266-2779; e-mail slo@mail.mcw.edu.

Received March 31, 2003; accepted October 9, 2003.

Previously published online at DOI: 10.1373/clinchem.2003.019216

⁴ Nonstandard abbreviations: BSA, bovine serum albumin; UBIL, unconjugated bilirubin; CAP, College of American Pathologists; TBIL, Vitros total bilirubin method; NB, Neonatal Bilirubin (Survey); C, Chemistry (Survey); DTB, ditaur bilirubin; NBIL, Vitros neonatal bilirubin method (sum of Bu + Bc); Bu, unconjugated bilirubin; and Bc, conjugated bilirubin.

mately three times greater than that of all Hitachi participants in the Survey (CV ~3.5%). Interestingly, the CVs for all instruments combined in the CAP (CAP-All Instruments) ranged from 7.9% to 8.2%, lower than those for the Vitros and Hitachi reported by Vreman et al. (3).

An editorial that accompanied the report by Vreman et al. (3) attributed the excessive variability of the results to lapses in quality assurance procedures, failure to calibrate instruments properly, and possible matrix effects (4).

Examination of the data of the 1995 CAP Survey (Table 1) shows that the mean total bilirubin values obtained with the Hitachi were 15–24 mg/L higher than the All-Instruments mean and that values from the Dimension were 23–34 mg/L higher. Mean values of the Vitros for specimens NB-10 and NB-11 were practically identical to the All-Instruments means, but the mean value for specimen C-96 (125 mg/L), which is the same as NB-11, was ~17% lower than the CAP All-Instruments mean (150.2 mg/L). This puzzling discrepancy was not investigated at that time.

To assess the performance of total bilirubin measurements in US laboratories, we analyzed specimens from the CAP Neonatal Bilirubin (NB) and Chemistry (C) Surveys (provided gratis by CAP for this study) by the reference method for total bilirubin (5) and compared the results with those reported by laboratories participating in the CAP Surveys. We also found an explanation for the different bilirubin values reported for the same specimen in the NB and C Surveys by the Vitros users.

Materials and Methods

Specimens NB-06, NB-12, NB-13, NB-14, and C-98 were from the 2001 CAP Surveys, whereas C-96, NB-01, and NB-02 were from the 2002 CAP Surveys. Specimen NB-13 is the same as C-98, and specimen NB-02 is the same as C-96. The Survey specimens were shipped to participants in August (NB-06), October (C-98), November (NB-12, NB-13, and NB-14), February (C-96), and April (NB-01 and NB-02). According to information provided by the proficiency material manufacturer, all Survey specimens were bovine serum with added UBIL and ditaurobilirubin

(DTB), certain “stabilizers”, and antimicrobial agents. The specimens were kept at -20°C and shipped frozen to participants. The bilirubin concentration in these specimens reportedly remains stable for at least 1 year.

In summarizing the results of the CAP surveys, we grouped similar instruments and methods from the same manufacturer into a single category (see Table 2). For example, included under the heading “Roche Hitachi” are Hitachi 911 and 917 and Roche Modular.

Solutions of UBIL and DTB in human serum (obtained after a 12-h fast from two of the authors) were prepared as described by Doumas and coworkers (5, 6). These solutions were dispensed in 1-mL cryogenic vials (Nalgene) and kept at -70°C ; the bilirubin concentrations in these solutions remained unchanged for at least 9 months (UBIL) and 16 months (DTB). These solutions as well as UBIL solutions in BSA were analyzed by the reference method and by the Vitros TBIL and Neonatal Bilirubin (NBIL) methods at Children’s Hospital of Wisconsin. The NBIL result is the sum of the unconjugated bilirubin (Bu) and the bilirubin sugar conjugates (Bc), each measured separately by the BuBc slide.

CAP Survey specimens were analyzed in duplicate on 2 days in three laboratories: the Reference Standards Laboratory of the Children’s Hospital of Wisconsin, the Wisconsin State Laboratory of Hygiene, and the Reference Laboratory of Ortho Clinical Diagnostics.

The 2001 and 2002 CAP Survey specimens were analyzed by the reference laboratories in March and July 2002, respectively.

Results

INTERLABORATORY AGREEMENT AND DIFFERENCES OF RESULTS BETWEEN THE REFERENCE AND PARTICIPANTS’ METHODS

Summarized in Table 2 are results from all methods and from four analyzers that provided 75% of the data for the NB Surveys and 83% of the data for the C Surveys. For the NB specimens, three of the four analyzers show very good precision, whereas the Hitachi was somewhat less precise. The interlaboratory CVs for the reference method were quite low, demonstrating the reliability and transferability of the method. Mean bilirubin values from the Hitachi and Synchron analyzers showed small differences from the reference method, whereas the differences for the Dade Dimension and Vitros were, respectively, large and variable. For the most part, the variability among laboratories (All-Instruments CVs) was smaller than in the 1995 Survey.

For three of the four analyzers (Dimension, Hitachi, and Synchron), CVs and differences from the reference method for specimens C-96 and C-98 were, as expected, essentially the same as those for NB-02 and NB-13 because these pairs of specimens differ only in their labeling. However, the means of these specimen pairs were highly divergent when measured by the Vitros users. The Vitros means for the NB specimens were higher than the values

Table 1. Total bilirubin results from selected instruments reported in the 1995 CAP Surveys.

Specimen	Reference method	CAP, all instruments	Dade Dimension	Roche Hitachi	Beckman Synchron	Ortho Vitros
NB-10						
Mean, mg/L		189.4	223.8	213.9	193.0	189.3
CV, %		8.1	2.4	3.7	3.5	2.7
NB-11 ^a						
Mean, mg/L		145.0	169.3	160.5	145.2	148.4
CV, %		7.9	1.7	3.4	2.0	3.2
C-96 ^a						
Mean, mg/L	143.0	150.2	173.4	165.3	148.0	125.0
CV, %		12.9	2.4	3.6	3.0	8.8

^a C-96 = NB-11.

Table 2. CAP Survey specimens 2001 and 2002: Precision of reference and selected bilirubin methods and differences^a between means of selected and reference methods.

CAP specimen ^b	Reference method ^c		CAP, all instruments (n _{NB} = 1674) ^d (n _C = 4920)		Dade Dimension (n _{NB} = 300) (n _C = 1176)		Roche Hitachi (n _{NB} = 122) (n _C = 602)		Beckman Synchron (n _{NB} = 262) (n _C = 1038)		Ortho Vitros (n _{NB} = 590) (n _C = 1202)	
	Mean, mg/L	CV, %	CV, %	Difference, %	CV, %	Difference, %	CV, %	Difference, %	CV, %	Difference, %	CV, %	Difference, %
NB-06	177.1	1.5	5.6	6.6	2.2	9.7	4.5	0.0	2.4	2.8	2.0	10.9
NB-12	90.1	2.0	5.0	6.4	2.0	9.1	3.9	2.1	2.6	1.4	2.3	9.3
NB-14	18.4	2.2	8.8	4.9	3.6	6.5	4.9	-1.0	5.7	5.4	5.3	3.3
NB-13	177.6	1.1	4.7	5.7	2.0	10.0	3.5	4.2	2.4	2.6	1.9	5.6
C-98 (NB-13)	177.6	1.1	10.6	0.5	2.2	9.9	4.2	3.6	2.8	3.3	3.2	-16.0
NB-01	135.6	0.2	5.3	5.5	2.2	8.9	3.8	1.1	2.7	1.8	2.0	6.0
NB-02	178.8	1.6	5.4	5.3	2.0	7.6	4.4	-2.2	2.5	1.3	2.2	8.7
C-96 (NB-02)	179.6	0.7	12.1	-3.1	2.1	7.6	5.0	-1.8	2.5	1.3	3.6	-21.6

^a Difference = mean value of selected method minus mean value of reference method expressed as percentage of the reference method value.

^b NB-06, NB-12, NB-14, NB-13, and C-98 from 2001 surveys; NB-01, NB-02, and C-96 from 2002 surveys.

^c Mean of six determinations from three laboratories.

^d n_{NB} = mean number of laboratories participating in the NB surveys; n_C = mean number of laboratories participating in the C surveys.

by the reference method, whereas for the C specimens, they were lower. These results are similar to those of the 1995 Survey (see Table 1).

Because of the large difference (9.3 mg/L) between the mean total bilirubin values for C-98 and NB-13, we asked the Data Processing Section of the CAP to recalculate the mean value for C-98 after excluding the values reported by the Vitros users. The recalculated mean for C-98 (188.4 mg/L) was almost identical to that of NB-13 (187.7 mg/L), indicating that the difference in means of C-98 and NB-13 was attributable to the results reported by the Vitros users, who presumably analyzed C-98 by the Vitros TBIL slide and NB-13 by the BuBc slide. Furthermore, after exclusion of the Vitros results, the original CV for C-98 (10.6%) was reduced to 5.6%, which is similar to the CV for NB-13 (4.7%). We also observed a large difference in CV for the identical specimens C-96 (12.1%) and NB-02 (5.4%). A plausible explanation for the large differences in the CV values is that in the NB Surveys, Vitros users reported results obtained by the BuBc slide, whereas in the C Survey both methods (BuBc and TBIL slides) were used by the participants. The Vitros slides were developed to measure Bu and bilirubin glucuronides in human serum, not in bovine serum containing a Bc surrogate, i.e., DTB.

CAP NB-SURVEY SPECIMENS ARE INAPPROPRIATE FOR EVALUATING THE VITROS METHOD FOR NEONATAL BILIRUBIN

We investigated the differences in total bilirubin values reported by the Vitros users for the two pairs of identical specimens NB-13/C-98 and NB-02/C-96. We analyzed the Survey specimens on the Vitros 700XR by the TBIL and BuBc slides. The results were similar to those obtained by the Vitros users in the Surveys, i.e., NBIL (Bu + Bc) values obtained by the Vitros BuBc slide exceeded, for every specimen, the values by the TBIL slide. At a low concentration (17 mg/L), the difference [(Bu + Bc) -

TBIL] was small (2 mg/L), whereas at a high bilirubin concentration (146 mg/L), the difference was large (55 mg/L). The Vitros NBIL method (BuBc slide) measures the bilirubin fractions by direct spectrophotometry, whereas total bilirubin is measured by a diazo method. It appears that the CAP Survey specimens exhibit matrix interference in the Vitros TBIL and NBIL methods.

We tried to identify the cause of the difference between the Vitros TBIL and the NBIL values. Knowing that the CAP Survey specimens consist of bovine serum, rather than human serum, enriched with UBIL and DTB, we measured total bilirubin in solutions of UBIL and DTB in fresh human serum by the reference method and by the two Vitros methods. For the high-UBIL solution, the Vitros TBIL and NBIL values were higher than that of the reference method by 4.8% and 8.8%, respectively (Table 3). These differences could be attributable to inadequate calibration of the two Vitros methods. Although the reference method values derived from analyses performed over a 9- to 12-month period, both differences were statistically significant because the long-term imprecision of the reference method is low.

For the high-DTB solution, the Vitros TBIL value was 14% lower than that of the reference method, whereas the NBIL value was higher by ~19%. Thus, the concentration

Table 3. Analysis of bilirubin (mg/L) solutions in human serum by the reference method and the Vitros TBIL and BuBc methods.

Specimen	Reference method			Bilirubin, mg/L			
	Mean, mg/L	n ^a	CV, %	TBIL	Bu	Bc	Bu + Bc
Low UBIL	45.1	44	0.8	46	48	0	48
High UBIL	174.6	46	0.6	183	190	0	190
Low DTB	26.7	50	1.1	25	8	22	30
High DTB	161.4	48	0.9	139	17	176	193

^a Number of determinations.

of DTB is underestimated by the Vitros TBIL method and overestimated by the NBIL method even when the matrix is human serum. The presence of DTB may not be the only problem of the NB-Survey specimens. A second "off-fender" is most likely the protein base (bovine serum) of the specimens. Analysis of a specimen consisting of 102 mg/L UBIL in BSA (used because bovine serum was not available to us) with the Vitros 700 XR gave TBIL and NBIL values of 105 and 141 mg/L, respectively. Thus, specimens containing DTB and BSA or, most likely, bovine serum are inappropriate for evaluating the accuracy of either the TBIL or NBIL method on the Vitros. This finding is especially important in interpreting bilirubin data reported in NB Surveys by the Vitros users, who produce 33% of all results. An appropriate specimen for any bilirubin method would be UBIL in human serum.

The effects of protein matrix on the reactivity of DTB and on its absorbance spectrum have been published previously (6). With the reference method, DTB and UBIL have similar reactivities in human serum, human serum albumin, bovine serum, and BSA. Furthermore, the molar absorptivity of the DTB alkaline azopigment at 598 nm is the same as that of UBIL. The underestimation of DTB with the Vitros TBIL slide could be explained if the azopigments deriving from UBIL and DTB have different molar absorptivities. This possibility exists because the diazo reagent of the TBIL slide is different from diazotized sulfanilic acid.

COMPARISON OF IMPRECISION IN 1995 AND 2001

The precision data for the CAP NB Surveys from 1995 and 2001 are summarized and compared in Table 4. Specimen NB-14 in the CAP 2001 survey is omitted from this summary because of its low total bilirubin concentration (18.4 mg/L). The range of CVs for the four selected instruments is good and clinically acceptable. The precision of these instruments (all show CVs <5%) has not changed significantly since 1995, suggesting that reagents and calibrators for total bilirubin assays are being manufactured with consistent quality. The improved precision of the CAP All Instruments in 2001 suggests that some of the assays with poorer precision in 1995 had been improved or discontinued in 2001.

Table 4. Range of CVs for selected instruments in CAP NB Surveys of 1995 and 2001/2002.

	CV, %	
	1995	2001/2002
CAP, all instruments	7.9–8.3	4.7–5.6
Dimension	1.4–2.4	2.0–2.2
Hitachi	2.6–3.7	3.5–4.5
Synchron	2.9–3.5	2.4–2.7
Vitros ^a	2.7–3.8	1.9–2.3

^a BuBc slide.

Discussion

Because DTB is almost invariably present in commercial controls and calibrators (total bilirubin and direct bilirubin assays are calibrated with a single solution), manufacturers of bilirubin assays based on direct spectrophotometry need to be aware of the dependence of the DTB spectrum on the protein matrix and pH. Given identical DTB concentrations, the spectra of DTB at pH 7.4 (phosphate buffer) in the previously mentioned protein matrices are different (6). In BSA and bovine serum, the spectrum of DTB is biphasic. In human serum albumin and human serum, the DTB spectrum has only one absorbance peak, at 460 nm. In water, the DTB spectrum peak shifts to 450 nm. Changing the pH to 8.5 (Tris buffer) causes the spectra to coalesce into a single spectrum with a peak at ~460 nm, but with different molar absorptivities. Furthermore, the molar absorptivities of UBIL are different in human serum and BSA (7, 8). Thus, calibrators consisting of UBIL in BSA (or, presumably, bovine serum) may not have the same values for diazo methods as for direct spectrophotometry. It is possible that the large differences in bilirubin values obtained by the BuBc slide of the Vitros are attributable to the differences in the absorption spectra and molar absorptivity values of DTB and UBIL between bovine serum and human serum.

Achieving accuracy in total bilirubin assays should not be difficult. A Standard Reference Material, SRM 916 Bilirubin, is available from the NIST, and a reference method for total bilirubin (5) has been credentialed by the National Reference System for Clinical Laboratories (9). Field methods should be traceable to the reference method and field calibrators to SRM 916. Table 5 shows what can be accomplished by use of the reference method and SRM 916. The molar absorptivities of SRM 916 and its azopigment have been determined twice, in 1989 and 2001, in US and European laboratories. There is a remarkable agreement in the absorptivity values determined on two occasions 12 years apart. The overall uncertainty of the mean values of molar absorptivity (SD of a single

Table 5. Corrected molar absorptivities of NIST Bilirubin (SRM 916a) and its neutral and alkaline azopigments.

	$\epsilon, \text{L} \cdot \text{mol}^{-1} \cdot \text{cm}^{-1}$			
	Caffeine reagent		Reference method	
	432 nm ^a	457 nm ^a	530 nm ^b	598 nm ^b
1989				
Weighted mean	50 060	48 890	56 660	76 490
SD of single measurement	580	570	1030	610
2001				
Weighted mean	50 090	49 240	57 090	76 640
SD of single measurement	850	440	730	600

^a Bilirubin.

^b Azopigments.

measurement) is very small, with CVs for the reference method <1% (10, 11).

In the preparation of bilirubin calibrators, the amount of bilirubin weighed is small, on the order of milligrams. To ascertain that bilirubin calibrators are properly prepared, they should be analyzed by the reference method and the molar absorptivity at 598 nm should be calculated. We recommend acceptability limits of 3 SD above and below the mean value. An additional criterion for acceptability of calibrators is the molar absorptivity of bilirubin in the caffeine reagent. The advantage of this approach, proposed by Vink et al. (12), is that the molar absorptivities of bilirubin at 432 and 457 nm are identical in the three most common protein matrices, i.e., human serum, human serum albumin, and BSA. The primary calibrator for total bilirubin measurement should be SRM 916 (UBIL) in human serum. For methods based on the Jendrassik–Gróf principle, calibrators made in human serum, bovine serum, human serum albumin, and BSA are equivalent.

The highest total bilirubin calibrator should be set at 250 mg/L because this is a clinical decision cutoff for the management of hyperbilirubinemia (1). In a healthy, term newborn 48 h or older, a serum total bilirubin \geq 250 mg/L will trigger an exchange transfusion if intensive phototherapy fails. A concentration of 250 mg/L is within the linear range of the total bilirubin assays on most major instruments. The Beckman CX and LX Synchron instruments have linear ranges up to 300 mg/L, the Dade Dimension to 250 mg/L, the Roche Hitachi to 300 mg/L, and the Ortho Clinical Diagnostics Vitros to 270 mg/L.

The introduction of proficiency testing specimens based on human serum would provide meaningful information regarding the accuracy and precision of total bilirubin assays. Recently, the CAP has introduced a limited number of human-serum-based specimens in the NB and C Surveys. If this trial is successful and becomes routine, the ability to grade proficiency testing participants for accuracy, using target values determined by the reference method, could become a valuable contribution toward improving the quality of bilirubin assays, which is essential for the management of neonatal hyperbilirubinemia.

We thank Bernadine Jendrzeczek and Laura Hubbard (Children's Hospital of Wisconsin), Glenn Ehlers and Eloise Casto (Ortho Clinical Diagnostics), and Barbara Hill and Michelle Gozda (Wisconsin State Laboratory of Hygiene) for performing direct bilirubin and total bilirubin assays, and the CAP for analyzing data and providing the Survey specimens.

References

1. American Academy of Pediatrics, Provisional Committee for Quality Improvement and Subcommittee on Hyperbilirubinemia. Practice parameter: management of hyperbilirubinemia in the healthy term newborn. *Pediatrics* 1994;94:558–65.
2. Schreiner RL, Glick MR. Interlaboratory bilirubin variability. *Pediatrics* 1982;69:277–81.
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. Doumas BT, Eckfeldt JH. Errors in measurement of total bilirubin: a perennial problem [Editorial]. *Clin Chem* 1996;42:845–8.
5. Doumas BT, Kwok-Cheung PP, Perry BW, Jendrzeczek 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. Doumas BT, Wu T-W, Kwok-Cheung PP, Jendrzeczek B. Chemical nature of a synthetic bilirubin conjugate and its reactivities in the total and direct reactions by the Jendrassik–Gróf method. *Clin Chem* 1985;31:1677–82.
7. Wennberg RP, Cowger ML. Spectral characteristics of bilirubin-bovine albumin complexes. *Clin Chim Acta* 1973;43:55–64.
8. Doumas BT, Perry BW, Sasse EA. The effect of protein base on the molar absorptivity and the absorption spectra of unconjugated bilirubin [Abstract]. *Clin Chem* 1973;19:647.
9. National Committee for Clinical Laboratory Standards. Bilirubin; approved summary of methods and materials credentialed by the NRSC Council; approved guideline. NCCLS Document RS6-A. Wayne, PA: NCCLS, 1988.
10. Doumas BT, Perry BW, McComb RB, Kessner A, Vader HL, Vink KLJ, et al. Molar absorptivities of bilirubin (NIST 916a) and its neutral and alkaline azopigments. *Clin Chem* 1990;36:1698–701.
11. National Institute of Standards and Technology. Certificate of analysis, Standard Reference Material 916a, bilirubin. Gaithersburg, MD: NIST, 2001.
12. Vink KLJ, Schuurman W, van Barsewinkel R. Use of caffeine reagent in direct spectrophotometry of bilirubin. *Clin Chem* 1986;32:1389–93.