JOURNAL OF CLINICAL MICROBIOLOGY, Dec. 2001, p. 4407–4412 Vol. 39, No. 12
0095-1137/01/$04.00+0 DOI: 10.1128/JCM.39.12.4407–4412.2001

European Proficiency Testing Program for Molecular Detection and Quantitation of Hepatitis B Virus DNA

ELIZABETH VALENTINE-THON,[1]* ANTON M. VAN LOON,[2] JURJEN SCHIRM,[3] JIM REID,[4] PAUL E. KLAPPER,[5] AND GRAHAM M. CLEATOR[5]

Department of Molecular Biology, Laboratory Dr. Schiwara and Partners, 28357 Bremen, Germany[1]; Department of Virology, University Medical Center Utrecht, Utrecht,[2] and Department of Virology, Regional Public Health Laboratory, Groningen,[3] The Netherlands; and Shield Diagnostics Limited, Dundee,[4] and Department of Virology, Manchester Royal Infirmary, Manchester,[5] United Kingdom

Received 30 April 2001/Returned for modification 14 August 2001/Accepted 27 September 2001

External quality control of hepatitis B virus (HBV) DNA detection remains an important issue. This study reports and compares the results obtained from two different proficiency panels for both the qualitative and quantitative assessment of HBV DNA. The panels were designed by the European Union Quality Control Concerted Action, prepared by Boston Biomedica, Inc., and distributed in May 1999 (panel 1) and February 2000 (panel 2). Each contained two negative samples and six positive samples with 10^3 to 10^7 copies/ml (panel 1) or 10^3 to 2×10^6 copies of HBV DNA per ml (panel 2). For panel 1, 42 laboratories submitted 20 qualitative (all in-house PCRs) and 37 quantitative (87% commercial assays) data sets. For panel 2, 51 laboratories submitted 25 qualitative (all in-house PCRs) and 47 quantitative (94% commercial assays) data sets. Five data sets (8.8%) in panel 1 and two data sets (2.8%) in panel 2 contained totals of six and two false-positives, respectively, corresponding to false-positive result rates of 5.3% for panel 1 and 1.4% for panel 2. The false-negative result rates of 10.5% for panel 1 and 17.4% for panel 2 were dependent on the detection levels of the assays employed as well as panel composition. In the qualitative analysis of all data sets, 47.4% (panel 1) and 51.4% (panel 2) had all samples correct. An adequate or better score (all correct or only the weak-positive sample missed) was obtained with 77.2% of the panel 1 samples and 68.1% of the panel 2 samples. In the quantitative analysis, 57.1% (panel 1) and 42.6% (panel 2) of the data sets achieved an adequate or better score (positive results within the acceptable range of the geometric mean ± 0.5 $\log_{10}$ of all positive results). These results demonstrate that while the qualitative performance of HBV detection has considerably improved compared to that of a previously published HBV proficiency study, the detection levels of many commercial quantitative assays are still too high to allow adequate quantitation of all relevant clinical samples.

Direct detection and quantitation of hepatitis B virus (HBV) DNA in plasma or serum are now used routinely to evaluate viremia in HBV-infected persons, to identify infectious chronic carriers, and to predict and monitor the efficacy of antiviral therapy (2, 8, 11). Since the early 1980s, a variety of molecular detection and quantitation methods have been developed, including dot and slot blot hybridization with radioactive and nonradioactive DNA probes (19–21), chemiluminescent detection of HBV DNA-RNA hybrids (14), PCR amplification of HBV DNA followed by hybridization to probes bound to a microwell plate (10, 12, 22) or magnetic beads (13), branched DNA (bDNA) signal amplification of an HBV DNA-DNA hybrid (7), transcription-mediated amplification (9), and fluorescent real-time detection of amplified HBV DNA (1). Each method, calibrated uniquely, exhibits its own sensitivity, specificity, and dynamic range. Standardization is ongoing (5, 6).

To assess the relative value of these methods in detecting and quantitating HBV DNA, international proficiency studies with well-characterized, simulated clinical samples would be required. In the first and only such study published to date (17), 39 laboratories analyzed 22 samples, including 12 undiluted samples with and without HBV DNA. (The lowest positive sample contained 3.5 pg/ml, or approximately 980,000 copies/ml.) Only 27.9% of the data sets had all 12 samples correct, and 34.9% showed false-positive results. Clearly, a majority of the participating laboratories had problems with both sensitivity and specificity.

The present report describes two recent HBV proficiency panels (lowest viral load of 1,000 copies/ml) designed by the European Union Concerted Action on Quality Control (EU QCCA) of Nucleic Acid Amplification in Diagnostic Virology and prepared by Boston Biomedica, Inc. (BBI; West Bridgewater, Mass.). The results obtained with these panels demonstrate that while the qualitative detection of HBV DNA has significantly improved, the detection levels of many commercial quantitative assays are still too high to allow adequate quantitation of the clinical samples seen in routine diagnostic laboratories.

MATERIALS AND METHODS

Panels. (i) Preparation. Panels were prepared by BBI from human plasma containing HBV DNA of subtype *ad* or *ay* by appropriate dilution in sterile filtered defibrinated plasma (Basematrix) with 0.09% sodium azide as preservative in accordance with the ISO 9001 Quality System Standards and the 21CFR 820 "*Good Manufacturing Practice for Medical Devices: General*." Plasma units were obtained from Food and Drug Administration-licensed facilities that comply with the applicable federal regulations (21CFR, part 600).

The pilot dilutions made for each sample were tested by reference laboratories designated by the EU QCCA and by the reference laboratories of selected

* Corresponding author. Mailing address: Department of Molecular Biology, Laboratory Dr. Schiwara and Partners, Haferwende 12, 28357 Bremen, Germany. Phone: 0049-421-20720. Fax: 0049-421-2072-167. E-mail: Elizabeth.Valentine-Thon@Schiwara.de.

TABLE 1. Methods employed in proficiency panels 1 and 2

Assay	Detection limit (copies/ml)	No. of data sets	
		Panel 1 (n = 57)	Panel 2 (n = 72)
Qualitative			
In-house PCR	30–10,000[a]	13	18
In-house nested PCR	50–1,000	7	7
Quantitative			
Roche Monitor	400	16	20
Digene HCSI	1,400,000	7	3
Digene HCSII Standard	141,500	6[b]	10
Digene HCSII Ultrasensitive	4,700		5
Bayer bDNA	700,000	2	5
BAG AcuGen	1,000	1	1
In-house nested PCR	50–100	3	0
In-house PCR	1,000–10,000[c]	2	3

[a] Upper range for panel 2 = 100,000 copies/ml.
[b] Includes both HCSII standard and HCSII Ultrasensitive.
[c] Detection limit for panel 2 = 100 copies/ml.

diagnostic manufacturers. After assessment of results and approval by the EU QCCA, samples were dispensed in 2.2-ml aliquots at the appropriate dilutions and stored at −70°C until shipment to the participants in May 1999 (panel 1) and February 2000 (panel 2).

(ii) Composition. Each panel consisted of eight coded samples. Six samples contained HBV DNA with approximate target levels of 10^3 to 10^7 copies/ml (panel 1) and 10^3 to 2×10^6 copies/ml (panel 2). Two samples contained no virus and served as negative controls. To evaluate interassay reproducibility, three identical samples were included in both panels: 2×10^6 copies/ml each for *ad* and *ay* and 2×10^5 copies/ml for *ad*. To assess a possible effect of HBV subtype, each panel contained two pairs of samples with identical viral loads, but different subtypes.

Participants. The panels were distributed on dry ice by courier service to 45 laboratories (panel 1) and 61 laboratories (panel 2). The recipient laboratories were asked to report the arrival and condition of the panel immediately by fax and to return the results as soon as possible, but within 6 weeks (panel 1) or 4 weeks (panel 2). A code number, known only to the Neutral Office, University of Manchester, Manchester, United Kingdom, identified each laboratory. Laboratories participating in both proficiency studies were assigned the same code for both panels. A questionnaire was also sent to obtain technical information on the procedures employed. To ensure confidentiality, all laboratories sent their results to the Neutral Office. The results were analyzed anonymously at the Department of Molecular Biology, Laboratory Dr. Schiwara and Partners, Bremen, Germany.

After the closing date for each panel, each participating laboratory was sent a certificate of participation, the code of the panel for individual performance assessment, and a written report summarizing all results.

Qualitative analysis. For qualitative analysis, the results from the quantitative data sets were converted to qualitative data (i.e., positive or negative) and considered together with the true qualitative data sets. To assess performance, the following scoring system was applied: 1 point was given for each correct result for the true-positive and true-negative samples. A point was deducted for each false-positive or false-negative result, with the exception of the weak-positive sample containing 10^3 copies/ml. Thus, the maximum score to be obtained was 8 points, which was qualified as "good." Scores of 7 and 6 points were considered "adequate" and "mediocre," respectively, while <6 points was considered "poor."

Quantitative analysis. For quantitative analysis, the overall geometric mean (GM) and standard deviations (SD) were calculated for each (positive) sample for all assays, as well as separately according to the method used. To assess performance, the following scoring system was utilized: 1 point was awarded for each viral load result that was within the range of ± 0.5 $\log_{10}$ of the overall GM of each sample. Scoring was for positive samples only. For all other results, i.e., each positive result outside this range, false-positive results, or negative results on positive samples, no point was given or deducted. The maximum score of 6 points was considered "good," 5 and 4 points were considered "adequate" and "mediocre," respectively, and <4 points was considered "poor."

RESULTS

Participants and methods. In panel 1, 42 laboratories from 19 countries submitted 20 qualitative (all in-house PCRs) and 37 quantitative (87% commercial assays) data sets. In panel 2, 51 laboratories from 18 countries submitted 25 qualitative (all in-house PCRs) and 47 quantitative (94% commercial assays) data sets. The methods utilized and the corresponding detection limits as reported by the participants are listed in Table 1.

Qualitative analysis. (i) Panel 1. Correct results for the two negative samples were reported in 52 of the 57 data sets (91.2%), and a false-positive result was obtained in 5 data sets, 1 of which contained two false-positive results (Table 2). Of these six false-positive results (6 of 114 negative samples = 5.3%), three were obtained with commercial kits (two by Digene Hybrid Capture Systems and one by Roche Monitor Assay). The weak-positive sample (10^3 copies/ml) was correctly reported positive in 30 data sets. The 27 data sets with false-negative results were obtained with the Digene Hybrid Capture Systems (13 of 13), an in-house PCR (7 of 15), an in-house nested PCR (4 of 10), the Roche Monitor Assay (1 of 16), the Bayer bDNA Assay (1 of 2), and the BAG AcuGen Test (1 of 1). Another low-positive sample (2×10^5 copies/ml) was correctly reported positive in 48 data sets. The nine false-negative

TABLE 2. Overall qualitative results[a]

No. of false qualitative results[b]									
Panel 1					Panel 2				
Subtype	Target level (copies/ml)	Qualitative (n = 20)	Quantitative (n = 37)	Total (n = 57)	Subtype	Target level (copies/ml)	Qualitative (n = 25)	Quantitative (n = 47)	Total (n = 72)
ad	10,000,000	0	0	0	*ad*	2,000,000	1	0	1
ay	10,000,000	0	0	0	*ay*	2,000,000	1	0	1
ad	2,000,000	0	0	0	*ad*	200,000	1	9	10
ay	2,000,000	0	0	0	*ay*	200,000	1	9	10
ad	200,000	2	7	9	*ay*	20,000	3	18	21
ad	1,000	9	18	27	*ay*	1,000	8	24	32
—	0	0	3	3	—	0	1	0	1
—	0	1	2	3	—	0	1	0	1

[a] Data for qualitative and quantitative data sets were combined for qualitative analysis.
[b] For panel 1, the percentage of false-positive results was 5.3%, and the percentage of false-negative results was 10.5% For panel 2, the percentage of false-positive results was 1.4%, and the percentage of false-negative results was 17.4%.

TABLE 3. Performance scores for qualitative results[a]

Performance score	No. of results[b]					
	Panel 1			Panel 2		
	Qualitative (n = 20)	Quantitative (n = 37)	Total (n = 57)	Qualitative (n = 25)	Quantitative (n = 47)	Total (n = 72)
Good, 8	10	17	27 (47)	15	22	37 (51)
Adequate, 7	7	10	17 (30)	5	7	12 (17)
Mediocre, 6	3	8	11 (19)	1	0	1 (1)
Poor, <6	0	2	2 (4)	4	18	22 (31)

[a] Data for qualitative and quantitative data sets were combined for qualitative analysis.
[b] For panel 1, the percentage of results with scores of good plus adequate was 77.2%. For panel 2, the percentage of results with scores of good plus adequate was 68.1%.
[c] Values in parentheses are percentages.

results with this sample were in data sets obtained with the Digene Hybrid Capture Systems (7 of 13) and an in-house PCR (2 of 15). All other positive samples ($\geq 2 \times 10^6$ copies/ml) were correctly identified as positive in all data sets. The overall false-negative result rate was therefore 10.5% [36/(6 × 57] (Table 2).

A total of 27 data sets (47.4%) obtained the maximum score of 8 points, 17 (29.8%) had a score of 7 points, 11 (19.3%) had a score of 6 points, and 2 (3.5%) had a score of <6 points (Table 3).

(ii) Panel 2. Correct results for the two negative samples were reported in 70 data sets. A false-positive result was obtained in two qualitative data sets (both in-house PCRs); none of the quantitative data sets contained false-positive results (Table 2). Thus, the false-positive result rate was 1.4% (2 of 144 negative samples).

The weak-positive sample (10^3 copies/ml) was correctly reported positive in 40 of the 72 data sets (55.6%). The 32 negative results were obtained with all five Bayer bDNA assays, all three Digene Hybrid Capture System I tests, most (9 of 10) Digene Hybrid Capture System II tests, the 1 BAG AcuGen Test, 7 of 21 in-house PCRs, 1 of 7 in-house nested PCRs, and 2 of 20 Roche Monitor Assay tests. For the other positive samples, the negative result rates decreased with increasing sample viral load. The overall false-negative result rate was therefore 17.4% [75/(6 × 72] (Table 2).

A total of 37 data sets (51.4%) obtained the maximum score of 8 points, 12 (16.7%) had a score of 7 points, 1 (1.4%) had a score of 6 points, and 22 (30.6%) had <6 points (Table 3).

Quantitative analysis. (i) Panel 1. Quantitative HBV data were reported in 37 data sets, whereby results from 2 data sets (1 in-house PCR, all results listed as $<8 \times 10^3$;1 in-house nested PCR, all positive results listed as $>10^6$) could not be included in the calculations. Most data sets (87%) were derived from commercial kits. The overall GM and SD were calculated for each (positive) sample from all assays (Table 4) as well as separately for the 16 Roche data sets, the 13 Digene data sets, and the 6 remaining data sets taken together (data not shown). The overall GM compared well with the target levels. For most samples, ≥89% of the positive results were within the range of GM ± 0.5 $\log_{10}$. The GM values obtained with the Roche and Digene assays were remarkably similar and likewise corresponded well to the target levels (except for the weak-positive sample, which could not be detected by the Digene assays), although the SD values for the Roche assay were consistently larger than those for the Digene assays (data not shown). The GM values for the six remaining data sets calculated together were consistently lower than the GM for the Roche and Digene assays and showed consistently larger SD values (data not shown).

Table 5 shows the performance in the various data sets in relation to the assay used. Altogether, 58% of the quantitative data sets obtained a score of "adequate" or better. This included 14 of 16 (87.5%) of those obtained with the Roche assay. On the other hand, one of the three "poor" scores was likewise obtained with this assay. The other two "poor" scores were obtained with in-house nested PCRs.

(ii) Panel 2. Quantitative HBV data were reported in 47 data sets, most (94%) derived from commercial kits. The overall GM and SD were calculated for each (positive) sample from all assays (Table 4), as well as separately according to the method used (data not shown). The overall GM corresponded well with the target levels. For most samples, ≥83% of the positive results were within the range of GM ± 0.5 $\log_{10}$. As in

TABLE 4. Overall quantitative results for positive samples

Panel 1				Panel 2			
Target level ($\log_{10}$)	Subtype	No. (%) of positive results within range of GM ± 0.5 $\log_{10}$/no. tested	GM ± SD $\log_{10}$ for all assays (n = 35)	Target level ($\log_{10}$)	Subtype	No. (%) of positive results within range of GM ± 0.5 $\log_{10}$/no. tested	GM ± SD $\log_{10}$ for all assays (n = 47)
7.0	*ad*	31/35 (89)	6.93 ± 0.56	6.3	*ad*	39/47 (83)	6.35 ± 0.38
7.0	*ay*	33/35 (94)	7.05 ± 0.35	6.3	*ay*	40/47 (85)	6.38 ± 0.44
6.3	*ad*	31/35 (89)	6.25 ± 0.53	5.3	*ad*	34/38 (90)	5.41 ± 0.31
6.3	*ay*	32/35 (91)	6.36 ± 0.36	5.3	*ay*	34/38 (90)	5.57 ± 0.33
5.3	*ad*	22/28 (79)	5.33 ± 0.50	4.3	*ay*	27/29 (93)	4.16 ± 0.24
3.0	*ad*	15/16 (94)	3.34 ± 0.22	3.0	*ay*	15/23 (65)	3.23 ± 0.71

TABLE 5. Performance scores for quantitative results

Performance score	No. of results[a]									
	Panel 1				Panel 2					
	Roche (n = 16)	Digene (n = 13)[b]	Other (n = 6)[c]	All (n = 35)[d]	Roche (n = 20)	Digene I (n = 3)	Digene II (n = 10)	Digene U (n = 5)	Other (n = 9)[e]	All (n = 47)[d]
Good, 6	10	0	0	10 (29)	9	0	0	1	0	10 (21)
Adequate, 5	4	6	0	10 (29)	5	0	0	4	1	10 (21)
Mediocre, 4	1	7	4	12 (34)	4	0	8	0	0	12 (26)
Poor, <4	1	0	2	3 (8)	2	3	2	0	8	15 (32)

[a] For panel 1, the percentage of results with scores of good plus adequate was 57%. For panel 2, the percentage of results with scores of good plus adequate was 42%.
[b] Digene I and Digene II considered together.
[c] In-house nested PCR (n = 2), in-house PCR (n = 1), Bayer bDNA (n = 2), and BAG AcuGen (n = 1).
[d] Values in parentheses are percentages.
[e] Bayer bDNA (n = 5), in-house PCR (n = 3), and BAG AcuGen (n = 1).

panel 1, the GM for the Roche and Digene assays were similar but with larger SD values for the Roche assays (data not shown). The GM for the remaining assays tended to be lower than for the Roche and Digene assays, with the largest SD values obtained with the in-house quantitative PCRs (data not shown).

Altogether, 42% of the quantitative data sets obtained scores of adequate or better (Table 5). This score was achieved by all 5 users of Digene's Hybrid Capture System II Ultrasensitive Assay, by 14 (70%) users of the Roche Assay, and by the 1 user of the BAG AcuGen Assay, but by none of the users of the other methods.

Reproducibility. Interpanel reproducibility could be evaluated from the results obtained with three samples represented in both panels (Table 6). For the *ay* subtype sample with 2×10^6 copies/ml, the qualitative detection rates were virtually the same: 100 and 99%, respectively. However, the percentage of positive results within the range of ± 0.5 $\log_{10}$ of the GM for that sample was higher in panel 1 (91%) than in panel 2 (85%). The results for the *ad* subtype sample with the same viral load showed the same pattern: nearly identical qualitative detection rates (100 and 99%, respectively), but a higher percentage of positive results within the range of ± 0.5 $\log_{10}$ of the GM in panel 1 (89%) compared to that in panel 2 (83%).

For the sample containing 2×10^5 copies/ml, the reverse situation was true. While the qualitative detection rates remained about the same for panels 1 and 2 (84 and 86%, respectively), the percentage of positive results within the range of ± 0.5 $\log_{10}$ of the GM was considerably lower in panel 1 (79%) than in panel 2 (90%).

No differences in the ability of the assays to detect (Table 2) or quantitate (Table 4) the two HBV subtypes *ad* and *ay* could be observed in either panel.

DISCUSSION

To assess the value of currently available methods as utilized by diagnostic laboratories for detecting and quantitating HBV DNA, international proficiency studies with well-characterized, simulated clinical samples are required. In the first such study performed in conjunction with the European Expert Group on Viral Hepatitis (17), 39 laboratories submitted 43 data sets for 22 samples, including 12 undiluted samples (7 positive, 5 negative) with viral loads from 3.5 pg/ml (approximately 980,000 copies/ml) to 222 pg/ml (approximately 62,160,000 copies/ml). Viral loads were determined by a single assay (Genostics liquid hybridization test from Abbott). All but one laboratory used the PCR technique; none used the uracil *N*-glycosylase system to prevent contamination from previously amplified products. Of the 43 data sets, 15 (35%) showed false-positive results for the five negative samples. Furthermore, despite the relatively high value of the weakest sample (980,000 copies/ml), 16 of 43 data sets (37%) showed false-negative results. Only 12 of 43 data sets (28%) had all 12 samples correct. No quantitative analysis of the data was performed. The first study, however, clearly demonstrated the

TABLE 6. Interpanel reproducibility, for three samples

Sample	Detection rate in no. detected/tested (%)			GM ± SD	% within range (GM ± 0.5 $\log_{10}$)
	Qualitative	Quantitative	All		
1 (2×10^6 copies/ml [6.3 $\log_{10}$]; *ay*)					
Panel 1	20/20 (100)	37/37 (100)	57/57 (100)	6.36 ± 0.36	91
Panel 2	24/25 (96)	47/47 (100	71/72 (99)	6.38 ± 0.44	85
2 (2×10^6 copies/ml [6.3 $\log_{10}$]; *ad*)					
Panel 1	20/20 (100)	37/37 (100)	57/57 (100)	6.25 ± 0.53	89
Panel 2	24/25 (96)	47/47 (100)	71/72 (99)	6.35 ± 0.38	83
3 (2×10^5 copies/ml [5.3 $\log_{10}$]; *ad*)					
Panel 1	18/20 (90)	30/37 (81)	48/57 (84)	5.33 ± 0.50	79
Panel 2	24/25 (96)	38/47 (81)	62/72 (86)	5.41 ± 0.31	90

large number of laboratories with sensitivity and specificity problems in detecting HBV DNA.

Similar problems with sensitivity and specificity were reported from the early proficiency studies for *Mycobacterium tuberculosis* (16) and for hepatitis C virus (HCV) (24) (29% false-positive results and 39% false-negative results on 10 undiluted samples—4 positive and 6 negative—in 31 data sets). In the second HCV proficiency study, performed 3 years later (4), little improvement was reported (21% false-positive results, but 63% false-negative results for 10 undiluted samples—again 4 positive and 6 negative—in 136 data sets).

Compared to the first HBV proficiency study (17), the present study shows considerable improvement both in panel conception and results obtained. First, the samples in both panels were well characterized, having been defined by the manufacturers of several quantitative methods (Roche, Bayer, and Digene) and by three reference laboratories utilizing various assays. The close approximation of the GM of all test results from the participating laboratories supports the accuracy of the assigned target viral load. Secondly, the sample with the lowest viral load in both panels contained only 1,000 copies of HBV DNA per ml, reflecting more accurately the viral loads from patients undergoing antiviral therapy. Third, this study involved a large number of participants (42 and 51 in panels 1 and 2, respectively) and data sets (57 and 72 in panels 1 and 2, respectively). In addition, the participating laboratories are now utilizing a variety of methods for detecting and quantitating HBV DNA in addition to PCR, allowing such methods to be assessed in an international proficiency panel. Fourth, the data were analyzed not only qualitatively, but also quantitatively, by using a simple algorithm permitting comparison of the seven different quantitative assays employed. Finally, the description of two independent proficiency panels permits interpanel reproducibility testing.

One of the most significant results derived from this study is the lowest false-positive rate reported to date for any large proficiency panel. Similar low false-positive rates have been found in the EU QCCA HCV RNA and human immunodeficiency virus proficiency studies carried out simultaneously with this study (J. Schirm, A. M. van Loon, E. Valentine-Thon, J. Reid, P. E. Klapper, and G. M. Cleator, submitted for publication; A. M. van Loon, J. Schirm, E. Valentine-Thon, J. Reid, P. E. Klapper, and G. M. Cleator, unpublished data) and in the EU QCCA enterovirus proficiency study carried out several months earlier (23). Indeed, the rate of 8.8% (5 of 57) false-positive data sets or 5.3% (6 of 114) false-positive results for panel 1 improved even further to 2.8% (2 of 72) false-positive data sets or 1.4% (2 of 144) false-positive results for panel 2. These low rates may simply reflect the greater expertise of the participating laboratories compared to several years ago. The improvement may also result from the increasing use of commercial kits, many of which contain contamination control enzyme systems. Indeed, 56% of data sets in panel 1 and 61% of data sets in panel 2 were obtained from commercial kits. On the other hand, the use of commercial kits does not safeguard against false-positive results: while the only false-positive results in panel 2 were obtained with in-house PCRs, 3 of 6 false-positive results in panel 1 were obtained with commercial kits. Finally, the lower false-positive rate in panel 2 compared to that in panel 1 may also represent a beneficial effect of participation in this proficiency program: of the four laboratories generating false-positive results in panel 1, three participated in panel 2 without producing false-positive results. Furthermore, the two laboratories generating false-positive results in panel 2 had not participated in panel 1.

In contrast to the low rate of false-positives in this study, the rate of negative results for (true) positive samples was remarkably high: for panel 1, 43.9% (25 of 57) of data sets, or 10.5% [36/(57 $\times$ 6] of results; for panel 2, 45.8% (33 of 72) of data sets, or 17.4% [75/(72 $\times$ 6] of results. Obviously, this high rate reflects the deliberate low viremic composition of the samples, in particular in panel 2 ($\leq 2 \times 10^6$ copies/ml) coupled with the high detection limits of some commercial kits still in use (i.e., Hybrid Capture System I with 1.4×10^6 copies/ml or Bayer bDNA with 0.7×10^6 copies/ml). (While the negative results obtained with low viremic samples by using such kits could arguably be excluded from the category "false-negative," we have chosen to include them in this category for the sake of simplicity.) As a result, for both qualitative and quantitative analyses, the percentage of data sets achieving scores of adequate or better decreased in panel 2 compared to that in panel 1. While it is recognized that such a panel composition "penalizes" users of commercial kits with high detection limits, the increasing clinical requirement for low-level dynamic ranges dictates this necessity. A potentially relevant assay is the newly introduced Cobas Amplicor HBV Monitor Test from Roche Diagnostics. However, while its low detection limit (200 copies/ml) is suitable for monitoring patients undergoing therapy, its upper dynamic range of only 200,000 copies/ml requires pretest dilution of HBsAg- and HBeAg-positive samples, concomitantly increasing costs (13, 15). Alternatively, transcription-mediated amplification followed by hybridization of two probes with different specific activities allows a broad detection range of 5×10^3 to 5×10^8 copies/ml (9). In contrast to Cobas Amplicor, however, this method contains no internal control for inhibition and is not yet automated.

In addition to low detection levels and a broad dynamic range, appropriate quantitation assays should be calibrated to internationally defined reference standards, as now available for HCV RNA (18). In the present study, the quantitative analysis showed variations ranging from 6 to 35% (percent positive results outside the defined range) in the actual copy numbers assigned to samples. Similar method-related deviations in quantitation have been reported by others (3, 5) In 1999, the Eurohep Pathobiology Group established two international reference plasma preparations, each containing approximately 2.6×10^9 copies/ml, thereafter defined as 10^9 Eurohep units (6). These Eurohep samples have been used for the evaluation of commercial kits (10, 14), and one of these may be the basis of a World Health Organization reference sample. In the present study, a panel 1 sample (*ay* subtype) with 10^7 copies/ml could also serve as a candidate reference standard, because the GM $\pm$ SD $\log_{10}$ for all 35 assays was 7.05, and 94% of the positive results with this sample were within the range of GM $\pm$ 0.5 $\log_{10}$.

Finally, both qualitative and quantitative assays must yield reproducible results. The interpanel reproducibility of the three samples represented in both panels was excellent: overall qualitative detection rates were nearly identical. However, for the two samples with 2×10^6 copies/ml, the percentage of

positive results within the range of ± 0.5 $\log_{10}$ of the GM for that sample was higher in panel 1 than in panel 2, while for the lower viremic sample (2×10^5 copies/ml), the opposite was true. Apparently, the increased use of commercial kits in panel 2 led to an increase in "within-range" detection rates in the lower dynamic range—at the expense, however, of a decreased within-range detection rate in the upper dynamic range.

In conclusion, this extensive proficiency study demonstrates considerable improvement in the qualitative performance of currently available HBV DNA assays. At the same time, however, the results emphasize the need for commercial viral load kits with sufficiently low detection levels to permit adequate HBV DNA detection and quantitation in clinical samples.

ACKNOWLEDGMENTS

The Blood-Borne Virus Working Party of the EU QCCA is grateful to all participating laboratories, reference laboratories (University Hospital Rotterdam, Department of Virology, Rotterdam, The Netherlands; University Medical Center Utrecht, Department of Virology, Utrecht, The Netherlands; and Roche Diagnostics, Basel, Switzerland) and manufacturers (Roche Diagnostics, Bayer Diagnostics, and Digene Diagnostics) for their contribution to this HBV DNA proficiency study.

The EU QCCA Program was supported with a grant from the EU Biomed 2 Program.

REFERENCES

1. **Abe, A., K. Inoue, T. Tanaka, J. Kato, N. Kajiyama, R. Kawaguchi, S. Tanaka, M. Yoshiba, and M. Kohara.** 1999. Quantitation of hepatitis B virus genomic DNA by real-time detection PCR. J. Clin. Microbiol. **37:**2899–2903.
2. **Beyer, A., J. Braner, H. W. Doerr, and B. Weber.** 1998. Quantitation of viral load: clinical relevance for human immunodeficiency virus, hepatitis B virus and hepatitis C virus infection. Intervirology **41:**24–34.
3. **Butterworth, L.-A., S. L. Prior, P. J. Buda, J. L. Faoagali, and W. G. E. Cooksley.** 1996. Comparison of four methods for quantitative measurement of hepatitis B viral DNA. J. Hepatol. **24:**686–691.
4. **Damen, M., H. T. Cuypers, H. L. Zaaijer, H. W. Reesink, W. P. Schaasberg, W. H. Gerlich, H. G. Niesters, and P. N. Lelie.** 1996. International collaborative study on the second EUROHEP HCV-RNA reference panel. J. Virol. Methods **58:**175–185.
5. **Gerlich, W. H., K. H. Heermann, R. Thomssen, and the Eurohep Group.** 1995. Quantitative assays for hepatitis B virus DNA: standardization and quality control. Viral Hep. **1:**53–57.
6. **Heermann, K. H., W. H. Gerlich, M. Chudy, S. Schaefer, R. Thomssen, and the Eurohep Pathobiology Group.** 1999. Quantitative detection of hepatitis B virus DNA in two international reference plasma preparations. J. Clin. Microbiol. **37:**68–73.
7. **Hendricks, D. A., B. J. Stowe, B. S. Hoo, J. L. Kolberg, B. D. Irvine, P. D. Neuwald, M. S. Urdea, and R. P. Perrillo.** 1995. Quantitation of HBV DNA in human serum using a branched DNA (bDNA) signal amplification assay. Am. J. Clin. Pathol. **104:**537–546.
8. **Jardi, R., M. Buti, F. Rodriquez-Frias, M. Cortina, R. Esteban, J. Guardia, and C. Pascual.** 1996. The value of quantitative detection of HBV-DNA amplified by PCR in the study of hepatitis B infection. J. Hepatol. **24:**680–685.
9. **Kamisango, K., C. Kamogawa, M. Sumi, S. Goto, A. Hirao, F. Gonzales, K. Yasuda, and S. Iino.** 1999. Quantitative detection of hepatitis B virus by transcription-mediated amplification and hybridization protection assay. J. Clin. Microbiol. **37:**310–314.
10. **Kessler, H. H., K. Pierer, E. Dragon, H. Lackner, B. Santner, D. Stünzner, E. Stelzl, B. Waitzl, and E. Marth.** 1998. Evaluation of a new assay for HBV DNA quantitation in patients with chronic hepatitis B. Clin. Diagn. Virol. **9:**37–43.
11. **Lok, A. S. F.** 1994. Treatment of chronic hepatitis B. J. Viral Hep. **1:**105–124.
12. **Mantero, G., A. Zonaro, A. Albertini, P. Bertolo, and D. Primi.** 1991. DNA enzyme immunoassay: general method for detecting products of polymerase chain reaction. Clin. Chem. **37:**422–429.
13. **Marin, I. J., M. Poljak, K. Seme, J. Meglič-Volkar, M. Matičič, G. Lešničar, and V. Brinovec.** 2001. Comparative evaluation of semiautomated COBAS AMPLICOR hepatitis B virus (HBV) MONITOR Test and manual microwell plate-based AMPLICOR HBV MONITOR Test. J. Clin. Microbiol. **39:**758–761.
14. **Niesters, H. G. M., M. Krajden, L. Cork, M. de Medina, M. Hill, E. Fries, and A. D. M. E. Osterhaus.** 2000. A multicenter study evaluation of the Digene Hybrid Capture II signal amplification technique for detection of hepatitis B virus DNA in serum samples and testing of EUROHEP standards. J. Clin. Microbiol. **38:**2150–2155.
15. **Noborg, U., A. Gusdal, E. K. Pisa, A. Hedrum, and M. Lindh.** 1999. Automated quantitative analysis of hepatitis B virus DNA by using the Cobas Amplicor HBV Monitor Test. J. Clin. Microbiol. **37:**2793–2797.
16. **Noordhoek, G. T., A. H. J. Kolk, G. Bjune, D. Catty, J. W. Dale, P. E. M. Fine, P. Godfrey-Faussett, S.-N. Cho, T. Shinnick, S. B. Svenson, S. Wilson, and J. D. A. van Embden.** 1994. Sensitivity and specificity of PCR for detection of *Mycobacterium tuberculosis*: a blind comparison study among seven laboratories. J. Clin. Microbiol. **32:**277–284.
17. **Quint, W. G. V., R. A. Heijtink, J. Schirm, W. H. Gerlich, and B. Niesters.** 1995. Reliability of methods for hepatitis B virus DNA detection. J. Clin. Microbiol. **33:**225–228.
18. **Saldanha, J., N. Lelie, and A. Heath.** 1999. Establishment of the first international standard for nucleic acid amplification technology (NAT) assays for HCV RNA. W.H.O. Collaborative Study Group. Vox Sang. **76:**149–158.
19. **Scotto, J., M. Hadchouel, C. Hery, J. Yvart, P. Tiollais, and C. Brechot.** 1983. Detection of hepatitis B virus DNA in serum by a single spot hybridization technique: comparison with results for other viral markers. Hepatology **3:**279–282.
20. **Valentine-Thon, E., J. Steinmann, and W. Arnold.** 1990. Evaluation of the commercially available HepProbe kit for detection of hepatitis B virus DNA in serum. J. Clin. Microbiol. **28:**39–42.
21. **Valentine-Thon, E., J. Steinmann, and W. Arnold.** 1991. Detection of hepatitis B virus DNA in serum with nucleic acid probes labelled with ^{32}P, biotin, alkaline phosphatase or sulphone. Mol. Cell. Probes **5:**299–305.
22. **Valentine-Thon, E.** 1995. Evaluation of SHARP Signal System for enzymatic detection of amplified hepatitis B virus DNA. J. Clin. Microbiol. **33:**477–480.
23. **Van Vliet, K., P. Muir, J. M. Echevarria, P. E. Klapper, G. M. Cleator, and A. M. van Loon.** 2001. Multicenter proficiency testing of nucleic acid amplification methods for the detection of enteroviruses. J. Clin. Microbiol. **39:**3390–3392.
24. **Zaaijer, H. L., H. T. M. Cuypers, H. W. Reesink, I. N. Winkel, G. Gerken, and P. N. Lelie.** 1993. Reliability of polymerase chain reaction for detection of hepatitis C virus. Lancet **341:**722–724.