

External Quality Assessment Program for Qualitative and Quantitative Detection of Hepatitis C Virus RNA in Diagnostic Virology

Jurjen Schirm,^{1*} Anton M. van Loon,² Elizabeth Valentine-Thon,³ Paul E. Klapper,^{4,5}
Jim Reid,⁶ and Graham M. Cleator⁴

Department of Virology, Regional Public Health Laboratory, Groningen,¹ and Department of Virology, University Medical Center Utrecht, Utrecht,² The Netherlands; Department of Molecular Biology, Laboratory Dr. Schiwarra and Partners, Bremen, Germany³; and Department of Virology, Manchester Royal Infirmary, Manchester,⁴ The Public Health Laboratory, Leeds,⁵ and Chimerabio Ltd., Dundee,⁶ United Kingdom

Received 4 September 2001/Returned for modification 24 November 2001/Accepted 23 May 2002

To assess the performance of laboratories in detecting and quantifying hepatitis C virus (HCV) RNA levels in HCV-infected patients, we distributed two proficiency panels for qualitative and quantitative HCV RNA testing. 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 panel consisted of two negative samples and six positive samples, with HCV RNA target levels from 200 to 500,000 copies/ml. Panel 1 had four samples with at least 50,000 copies/ml, and panel 2 had two samples with at least 50,000 copies/ml. Fifty-seven laboratories submitted 45 qualitative and 35 quantitative data sets on panel 1, and 81 laboratories submitted 75 qualitative and 48 quantitative data sets on panel 2. In both panels, about two-thirds of the qualitative data sets and >90% of the quantitative data sets were obtained with commercial assays. With each panel, two data sets gave one false-positive result, corresponding to false-positivity rates of 1.3% and 0.8% for panel 1 and panel 2, respectively. Samples containing at least 50,000 copies/ml were found positive in 97% and 99% of the cases with panel 1 and panel 2, respectively. In contrast, the positive samples containing $\leq 5,000$ copies/ml were reported positive in only 71% and 77% of the cases with panel 1 and panel 2, respectively. Adequate or better scores on qualitative results (all results correct or only the low-positive samples missed) were obtained in 84% (panel 1) and 80% (panel 2) of the data sets. In the analysis of quantitative results, 60% (panel 1) and 73% (panel 2) of the data sets obtained an adequate or better score ($\geq 80\%$ of the positive results within the range of the geometric mean $\pm 0.5 \log_{10}$). Our results indicate that considerable improvements in molecular detection and quantitation of HCV have been achieved, particularly through the use of commercial assays. However, the lowest detection levels of many assays are still too high, and further standardization is still needed. Finally, this study underlines the importance of proficiency panels for monitoring the quality of diagnostic laboratories.

Detection and quantitation of hepatitis C virus (HCV) RNA levels in plasma has become an essential part of the diagnosis and management of HCV-infected patients. Qualitative HCV RNA tests are used to identify acute HCV infections as well as chronic HCV carriers. Quantitation of HCV RNA is used to predict and monitor the efficacy of antiviral therapy (5, 6). In recent years, a variety of commercial and noncommercial test systems have been developed for this purpose, including competitive reverse transcription (RT)-PCR, noncompetitive RT-PCR, branched-DNA signal amplification, and real-time PCR (3, 4, 10, 12, 18). Each of these methods was calibrated with proprietary standards and exhibits its own sensitivity, specificity, and dynamic range. Reagents for standardization have only very recently been introduced (2, 11), although they have not yet been used extensively in practice.

Obviously, laboratories performing HCV RNA tests should report accurate and reliable results regardless of the type of assay used. One of the best ways to assess the performance of individual laboratories is to distribute proficiency panels and to

evaluate all the test results. Early proficiency studies on the detection of HCV RNA showed high percentages of laboratories with specificity and sensitivity problems (1, 19). Similar problems have been reported for the molecular detection of hepatitis B virus (HBV) (9) and *Mycobacterium tuberculosis* (8).

An external quality assessment program for the evaluation of currently employed nucleic acid amplification methods was established by the members of the European Union (EU) Quality Control Concerted Action of Nucleic Acid Amplification in Diagnostic Virology (QCCA). Between 1997 and 2000, proficiency panels were distributed for the detection of enterovirus RNA (16), herpes simplex virus DNA (L. Schloss, P. Cinque, G. M. Cleator, J.-E. Echevarria, K. I. Falk, P. E. Klapper, J. Schirm, B. F. Vestergaard, H. G. M. Niesters, T. Popow-Kraupp, W. G. V. Quint, A. M. van Loon, and A. Linde, unpublished data), cytomegalovirus DNA, *Chlamydia trachomatis* DNA (R. P. Verkooyen, G. T. Noordhoek, P. E. Klapper, J. Reid, J. Schirm, G. M. Cleator, and G. Hoddevik, unpublished data), human immunodeficiency virus RNA (A. M. van Loon, J. Schirm, E. Valentine-Thon, J. Reid, P. E. Klapper, and G. M. Cleator, unpublished data), HBV DNA (15), and HCV RNA.

The present report describes the qualitative and the quan-

* Corresponding author. Mailing address: Regional Public Health Laboratory, Van Ketwich Verschuurlaan 92, 9721 SW Groningen, The Netherlands. Phone: 31-50-5215 160. Fax: 31-50-5271 488. E-mail: schirmjsg@compuserve.com.

titative results obtained with the two QCCA HCV RNA panels distributed in 1999 and 2000. The results demonstrate that the quality of HCV RNA detection has clearly improved, particularly through the use of commercial assays. However, comparison of the quantitative results was hampered by the lack of standardization between different test types. Moreover, the lower detection limits of some of the quantitative assays used were too high for optimal monitoring of HCV-infected patients.

MATERIALS AND METHODS

Preparation of panels. The HCV RNA panels were designed by the EU QCCA Working Party on Blood Borne Viruses and produced by Boston Biomedica Inc. in accordance with the ISO 9001 Quality System Standards and 21 CFR 820 "Good Manufacturing Practice for Medical Devices: General." The panels were prepared from four different HCV RNA-positive human plasma samples containing HCV genotypes 1, 2, 3, and 4 by dilution with HCV-negative Basematrix (defibrinated plasma) to predefined target levels of HCV RNA. Basematrix had been sterilized by filtration and preserved by the addition of 0.09% sodium azide. HCV genotype assignment was based on testing with the InnoLipa assay and RNA sequencing. After preparation of pilot dilutions, the approximate target values were assigned by Boston Biomedica Inc. and confirmed by two major diagnostic manufacturers (Roche Diagnostics and Bayer Diagnostics). After approval of these testing results by the EU QCCA Working Party on Blood Borne Viruses, the final dilutions of the bulk stocks were prepared, dispensed in 1.1-ml portions per vial, and stored at -70°C until the time of shipment to the participants in May 1999 (panel 1) and February 2000 (panel 2). The final confirmation of the target loads was obtained through the geometric means of the results obtained by the participants (see below).

Composition. Each panel consisted of eight coded samples. Six samples contained HCV RNA with approximate target levels of 2×10^2 to 5×10^5 copies/ml. Two samples contained no virus and served as negative controls. To evaluate interassay reproducibility, four samples were included in both panels: 2×10^2 copies/ml (subtype 1), 5×10^3 copies/ml (subtype 1), 5×10^4 copies/ml (subtype 3), and 5×10^5 copies/ml (subtype 1). To assess a possible effect of HCV subtype, each panel contained pairs of samples with identical viral loads but different subtypes.

Distribution. All panels were distributed on dry ice by courier service from a central facility in Paris, France. Instructions for storage at -20°C or below and processing of the samples were enclosed, and a questionnaire was added in order to obtain technical information on the procedures employed by individual participants. The participating laboratories were asked to report receipt of the panel immediately by fax and to return the results as soon as possible but within 7 weeks to the Neutral Office, University of Manchester, Manchester, United Kingdom. If the panel did not arrive in good condition, a second shipment was made. A code number, known only to the Neutral Office, identified each laboratory. Laboratories participating in both proficiency studies were assigned the same code for both panels. Immediately after the closing dates, each participating laboratory was sent the code with target HCV RNA levels for individual performance assessment.

Analysis. All results were analyzed anonymously at the Department of Virology, Regional Public Health Laboratory, Groningen, The Netherlands. The overall evaluation for each panel was sent to the participants within a few months.

Analysis of qualitative results. The results from the quantitative data sets were converted to qualitative data (i.e., positive/negative) and considered together with the truly qualitative data sets. To assess the performances of individual participants, the following scoring system was applied. One point was given for each correct result. In addition, one point was deducted for each false-positive or false-negative result, with the exception of results on the relatively weak positive samples (target levels of $\leq 5 \times 10^3$ copies/ml). Thus, the maximum possible qualitative score was 8 points. Scores of 7 and 6 points were considered adequate and mediocre, respectively, while ≤ 5 points was considered poor.

Analysis of quantitative results. Although the HCV RNA target levels of the samples in each panel were expressed in copies per milliliter, some participants expressed their quantitative results in either genome equivalents per milliliter (only laboratories with the Quantiplex bDNA version 2.0 test) or international units per milliliter. According to a statement from the manufacturer of the bDNA test, 1 genome equivalent/ml equals 1 copy/ml (D. Hendricks, Bayer, 1999, personal communication). The four laboratories expressing their results in

TABLE 1. Methods used to detect HCV RNA^a

| Method ^b | No. (%) of data sets | |
|--------------------------------|----------------------|---------|
| | Panel 1 | Panel 2 |
| Qualitative | | |
| Roche Cobas/Amplicor | 7 (16) | 34 (45) |
| Roche Amplicor (manual) | 21 (47) | 16 (21) |
| Roche (details unknown) | | 2 (3) |
| In-house nested RT-PCR | 5 (11) | 13 (17) |
| In-house RT-PCR | 9 (20) | 7 (9) |
| Inno-Lipa ^d | 2 (4) | 2 (3) |
| Unknown method ^d | 1 (2) | 1 (1) |
| Total | 45 | 75 |
| Quantitative | | |
| Roche Cobas/Amplicor | 7 (20) | 24 (50) |
| Roche Amplicor (manual) | 16 (46) | 10 (21) |
| Bayer bDNA version 2.0 | 8 (23) | 11 (23) |
| BAG/AcuGen | 1 (3) | 1 (2) |
| In-house RT-PCR | 2 (6) | |
| In-house RT-PCR (Taqman) | 1 (3) | 1 (2) |
| In-house RT-PCR (Light Cycler) | | 1 (2) |
| Total | 35 | 48 |

^a Some laboratories submitted more than one data set.

^b Laboratories using Roche assays often provided no further details on the precise version of the test and whether they used manual or automated systems.

^c Additional HCV genotyping was performed for one of these data sets.

^d Additional HCV genotyping was performed on all of these data sets.

international units per milliliter (in the 2000 panel 2 only) all used a new version of the Cobas/Amplicor HCV Monitor version 2.0 assay. This system has different conversion factors from international units per milliliter to copies per milliliter for different lot numbers, in practice ranging between 0.6 and 3.8 (11). Unfortunately, the individual conversion factors were very difficult to obtain. For these reasons, we decided, for the sake of simplicity, to consider all quantitative data in this study to be expressed in copies per milliliter. Consequently, all calculations below are based on copies per milliliter.

For evaluation of the quantitative data sets, the copies per milliliter results were first converted to \log_{10} values, and then the overall geometric mean (GM) in \log_{10} copies/ml and the standard deviation (SD) were calculated for each (positive) sample from all reported quantitative positive results. To assess the performances of individual participants, we calculated what percentage of the reported positive results of each data set was within the acceptable range of $\text{GM} \pm 0.5 \log_{10}$. This range was chosen because viral load differences of $<0.5 \log_{10}$ are usually not considered clinically relevant. In addition, the SDs calculated for each sample in the present study were on average $0.45 \log_{10}$ (see below), which is very close to the chosen acceptable range. When at least 80% of the positive results reported by one quantitative data set were within the acceptable range, the quantitative performance was qualified as good (100%) or adequate (80 to 100%). Data sets with 60 to 80% acceptable quantitative results were considered mediocre, and $<60\%$ was qualified as poor.

RESULTS

Participants and methods. Panel 1 (1999) was tested by 57 laboratories from 20, mainly European, countries submitting 45 qualitative and 35 quantitative data sets. Panel 2 (2000) was tested by 81 laboratories from 22 countries submitting 75 qualitative and 48 quantitative data sets. With both panels, more than 90% of the participating laboratories were hospital laboratories or other diagnostic laboratories. Table 1 shows that about two-thirds of the qualitative data sets (64% and 72% for panels 1 and 2, respectively) and nearly all quantitative data sets (91% and 96%, respectively) were obtained with commercial assays. The three other quantitative data sets were obtained by in-house real-time PCR. In addition, three (panel 1) and four (panel 2) laboratories also performed supplementary HCV genotyping.

TABLE 2. Overall qualitative results

| Sample target concn (copies/ml) | Genotype | Panel 1 | | | Panel 2 | | |
|---------------------------------|----------|------------|---|-------------------------------|------------|---|-------------------------------|
| | | Sample no. | No. (%) of data sets with correct results | | Sample no. | No. (%) of data sets with correct results | |
| | | | Qualitative methods (n = 45) | Quantitative methods (n = 35) | | Qualitative methods (n = 75) | Quantitative methods (n = 48) |
| 500,000 | 1 | 1 | 44 (97.8) | 35 (100) | 3 | 75 (100) | 48 (100) |
| 50,000 | 1 | 4 | 44 (97.8) | 35 (100) | | | |
| 50,000 | 2 | 2 | 44 (97.8) | 33 (94.3) | | | |
| 50,000 | 3 | 8 | 43 (95.6) | 33 (94.3) | 6 | 74 (98.7) | 46 (95.8) |
| 5,000 | 1 | 7 | 41 (91.1) | 28 (80.0) | 5 | 71 (94.7) | 37 (77.1) |
| 5,000 | 1 | | | | 8 | 69 (92.0) | 35 (72.9) |
| 5,000 | 4 | | | | 2 | 71 (94.7) | 35 (72.9) |
| 200 | 1 | 5 | 33 (73.3) | 12 (34.3) | 4 | 54 (72.0) | 5 (10.4) |
| 0 | | 3 | 45 (100) | 35 (100) | 1 | 74 (98.7) | 48 (100) |
| 0 | | 6 | 45 (100) | 33 (94.3) | 7 | 74 (98.7) | 48 (100) |

Laboratories that used commercial assays often gave no details on the precise version of the test and whether or not they used manual (such as Amplicor) or semiautomated (such as Cobas/Amplicor) systems. Also, information on the lower detection limits of the tests used was reported by no more than about half of the participants, whereas not all laboratories using one particular test system reported the same lower detection limit for that system. The reported lower detection levels varied from 10 to 5,000 copies/ml for the qualitative assays and from 100 to 200,000 copies/ml for the quantitative assays.

Analysis of qualitative results. (i) Panel 1. A total of 80 qualitative data sets were available for analysis, 45 truly qualitative data sets and 35 derived from quantitative data sets. One of the two negative samples was reported positive in two data sets (Table 2), both produced with the quantitative Bayer bDNA test version 2.0 (viral loads of 238,000 and 257,000 copies/ml). The other negative sample was reported negative in all data sets. Consequently, 2 of 160 (1.3%) of all tests performed on negative samples were false-positive (0% of the qualitative tests and 2.9% of the quantitative tests). The four high-positive samples (1, 2, 4, and 8) were correctly reported positive in 97% of all data sets. The sample with a target level of 5,000 copies/ml (sample 7) was missed in 11 data sets obtained by six of the eight laboratories with the quantitative bDNA method, one of three quantitative in-house PCR assays, and 4 of 14 noncommercial qualitative methods. The weakly positive sample (target level of 200 copies/ml) was correctly reported positive in 45 of 80 data sets. The 35 data sets with negative results on this sample were obtained with qualitative Roche assays (3 of 28), other qualitative methods (9 of 17), and most of the quantitative test systems: 14 of 23 Roche Monitor assays, 7 of 8 bDNA assays, and 2 of 4 quantitative in-house PCR assays.

A total of 42 data sets (52.5%) obtained the maximum qualitative performance score of 8 points (Table 3). These data sets included 33 of 45 (73%) of the data sets obtained with qualitative methods and 9 of 35 (26%) of the data sets obtained with quantitative methods. An additional 25 (31.2%) data sets had a score of 7 points, 6 data sets (7.5%) had a score of 6 points, and 7 data sets (8.8%) had a score of ≤5 points.

(ii) Panel 2. A total of 123 qualitative data sets were available for analysis: 75 truly qualitative data sets and 48 derived

from quantitative data sets. Each of the two negative samples was reported positive in one data set (Table 2); these two data sets were submitted by one laboratory with a qualitative Cobas/Amplicor kit and by another laboratory performing qualitative in-house PCR. Consequently, 2 of 246 (0.8%) tests performed on negative samples were false-positive (1.3% of the qualitative tests and 0% of the quantitative tests). High-positive samples 3 and 6 were correctly identified in 100% and 97.6% of all data sets, respectively. The sample with a target level of 50,000 copies/ml was missed in three data sets (one by qualitative in-house nested PCR, two by bDNA).

The three samples containing target levels of 5,000 IU/ml (2, 5, and 8) were correctly identified by 94.7%, 94.7%, and 92.0%, respectively, of the qualitative data sets and by 72.9%, 77.1%, and 72.9%, respectively, of the quantitative data sets. Most of the negative results reported for these three positive samples were obtained by the quantitative bDNA assay (32 of 33), qualitative in-house nested PCRs (12 of 39), and the quantitative Roche Monitor assays (4 of 102). Comparison of samples 2, 5, and 8 shows that there was little difference between the qualitative results for HCV genotypes 1 and 4. The weak-positive sample (target level of 200 copies/ml) was correctly reported positive in 59 of 123 data sets. The 64 data sets with negative results on this sample were obtained with qualitative Roche assays (5 of 50), other qualitative methods (16 of 23), and most of the quantitative test systems: 30 of 34 Roche Monitor assays and all bDNA assays (11 of 11) and quantitative in-house PCR assays (2 of 2).

Table 3 shows that a total of 52 data sets (42.2%) obtained the maximum qualitative performance score of 8 points. These data sets included 49 of 75 (65%) of the data sets obtained with

TABLE 3. Performance scores for qualitative results

| Performance (score) | No. (%) of data sets | | | |
|---------------------|----------------------|-----------------------|----------------------|-----------------------|
| | Panel 1 | | Panel 2 | |
| | Qualitative (n = 45) | Quantitative (n = 35) | Qualitative (n = 75) | Quantitative (n = 48) |
| Good (8) | 33 (73) | 9 (26) | 49 (65) | 3 (6) |
| Adequate (7) | 7 (16) | 18 (51) | 15 (20) | 31 (65) |
| Mediocre (6) | 2 (4) | 4 (11) | 5 (7) | 3 (6) |
| Poor (≤5) | 3 (7) | 4 (11) | 6 (8) | 11 (23) |

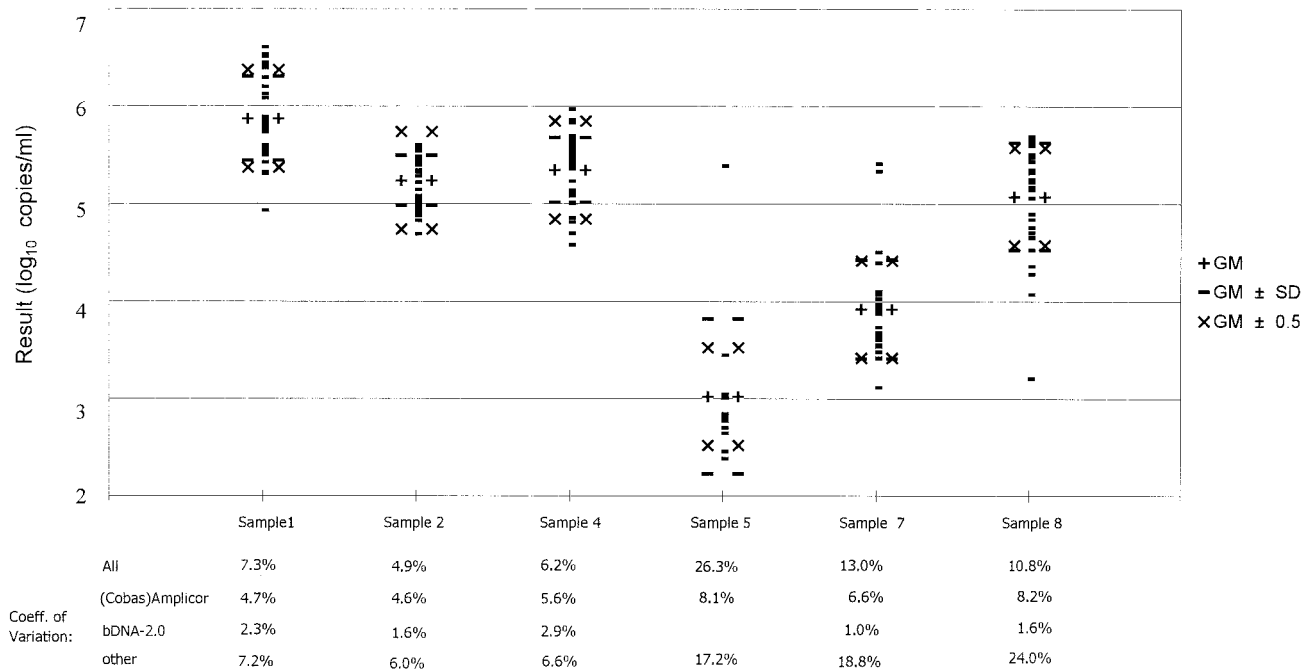


FIG. 1. GM (log₁₀), SD, and CV with various amplification methods for HCV RNA, panel 1.

qualitative methods and 3 of 48 (6%) of the qualitative data sets derived from quantitative results. An additional 46 (37.3%) data sets had a score of 7 points, 8 data sets (6.5%) had a score of 6 points, and 17 data sets (13.8%) had a score of ≤5 points.

HCV genotyping. Although HCV genotyping was not requested in our proficiency study, three and four laboratories performed HCV genotyping in 1999 and 2000, respectively. Altogether, 21 typing results were produced for the five samples containing HCV genotype 1. All results indicated the presence of genotype 1. Similarly, the seven typing results performed on the samples containing HCV genotype 3 were also correct. However, the genotyping of the samples containing HCV genotypes 2 and 4 was not entirely correct. For both samples, one laboratory incorrectly identified the virus as HCV genotype 1. This laboratory used an in-house multiplex RT-PCR method.

Analysis of quantitative results. (i) Panel 1. Quantitative HCV data were reported in 35 data sets, mostly (91%) obtained with commercial kits. The viral loads reported for the positive samples are summarized in Fig. 1 and Table 4. For each sample, the overall GM (log₁₀) and SD were calculated from the positive results obtained with all assays. The GMs for the different samples were all somewhat higher (0.17 to 0.65 log₁₀) than the target levels, especially for two of the samples with 50,000 copies/ml (2 and 4, both not used in panel 2). Table 4 shows that the percentage of positive results within the accepted range of GM ± 0.5 log₁₀ varied from 63% to 97%. When, in addition, the GMs were calculated for different methods separately (23 Roche data sets, 8 Bayer bDNA data sets, and the remaining 4 data sets taken together), most of the GMs were quite similar (± <0.5 log₁₀). However, for the strongest positive sample (number 1), relatively low viral loads were obtained with the Roche methods. In addition, the Bayer

TABLE 4. Summary of quantitative results^a

| Sample target concn (copies/ml) ^b | Genotype | Panel 1 | | Panel 2 | |
|--|----------|------------------------------|--|------------------------------|--|
| | | GM (log ₁₀) ± SD | No. (%) of positive results within the range GM (log ₁₀) ± 0.5/total | GM (log ₁₀) ± SD | No. (%) of positive results within the range GM (log ₁₀) ± 0.5/total |
| 500,000 (5.70) | 1 | 5.87 ± 0.43 | 22/35 (63) | 5.59 ± 0.34 | 40/48 (83) |
| 50,000 (4.70) | 1 | 5.35 ± 0.33 | 30/35 (86) | | |
| 50,000 (4.70) | 2 | 5.24 ± 0.26 | 32/33 (97) | | |
| 50,000 (4.70) | 3 | 5.08 ± 0.55 | 21/33 (64) | 4.97 ± 0.55 | 43/46 (72) |
| 5,000 (3.70) | 1 | 3.92 ± 0.51 | 23/28 (82) | 3.93 ± 0.50 | 31/37 (84) |
| 5,000 (3.70) | 1 | | | 3.78 ± 0.40 | 31/35 (89) |
| 5,000 (3.70) | 4 | | | 3.70 ± 0.35 | 31/35 (89) |
| 200 (2.70) | 1 | 3.02 ± 0.80 | 9/12 (75) | 2.89 ± 0.42 | 4/5 (80) |

^a Only positive results were included.

^b Values in parentheses are log₁₀ target concentration.

TABLE 5. Quantitative performances of HCV RNA data sets

| Performance level (% within range ^a) | No. (%) of data sets ^b | | | | | |
|---|-----------------------------------|-----------------------|------------------------|--------------------------|------------------------|------------------------|
| | Panel 1 | | | Panel 2 | | |
| | Roche assays (n = 23) | bDNA tests (n = 8) | Other tests (n = 4) | Roche assays (n = 34) | bDNA tests (n = 11) | Other tests (n = 3) |
| Good/adequate (≥80) | 19 (83) | 2 (25) | 0 (0) | 31 (91) | 4 (36) | 0 (0) |
| Mediocre/poor (<80) | 4 (17) | 6 (75) | 4 (100) | 3 (9) | 7 (64) | 3 (100) |

^a Percent of positive results within the range GM (log₁₀) ± 0.5.

^b The quantitative scores for the bDNA version 2.0 method were, for most laboratories, based on only two positive results. Other tests included in-house RT-PCR (n = 2), TaqMan RT-PCR (n = 1), and BAG AcuGen (n = 1) (panel 1) and TaqMan RT-PCR (n = 1), Light Cycler RT-PCR (n = 1), and BAG AcuGen (n = 1) (panel 2).

bDNA method gave consistently higher results (+0.62 to 1.63 log₁₀) with the relatively weak positive sample 7 (genotype 1) and with HCV genotype 3 (sample 8) (data not shown). Figure 1 also shows that the coefficients of variation (CV) for the different samples varied from 4.9% for one of the high-positive samples to 26.3% for the lowest positive sample. For all samples tested, the CVs were smaller with the bDNA test (1.9 to 2.9%) than with the Roche assays (4.6 to 8.2%).

Table 5 shows that 21 (60%) of the 35 quantitative data sets obtained a good or adequate quantitative performance score of ≥80%. These data sets included 83% of the data sets obtained with Roche assays, 25% of the data sets obtained with bDNA tests, and none of the data sets obtained with other methods. Consequently, 17% of the laboratories using Roche assays and the vast majority of all the other laboratories need to improve their performance of quantitative HCV RNA testing.

(ii) **Panel 2.** The HCV viral loads, as reported in 48 data sets, are summarized in Table 4. For each sample, the overall GM (log₁₀) and SD were calculated from the positive results obtained with all assays. This showed that the overall GMs for the different samples were all close to the target levels (difference of <0.30 log₁₀) and also, for the four samples used in both panels, close to the GMs obtained with panel 1 (see also Table 6). Table 4 also shows that the percentage of positive results within the accepted range of GM ± 0.5 log₁₀ varied between

72% and 89%. When, in addition, the GMs were calculated for the different test methods separately (34 Roche data sets, 11 Bayer bDNA data sets, and the remaining 3 data sets taken together), the GMs of the Roche tests were always much lower (0.58 to 1.8 log₁₀) than the GMs for the bDNA assays but much higher (up to 1.29 log₁₀) than the GMs for the three other data sets combined. The overall CV for the different samples varied from 6.0% for the highest positive samples to 14.5% for the weak-positive sample. For the two highest positive samples (3 and 6), the CVs were much smaller with the bDNA test (2.8 and 2.5%, respectively) than with the Roche assays (4.0 and 8.2%, respectively).

Table 5 shows that 35 (73%) of the 48 quantitative data sets obtained a good or adequate quantitative performance score of ≥80%. These data sets comprised 91% of the data sets obtained with Roche assays, including three of four data sets reporting in international units per milliliter, 36% of the data sets obtained with bDNA tests, and none of the data sets obtained with other methods. Consequently, 9% of the laboratories using Roche assays and the vast majority of all the other laboratories need to improve their performance of quantitative HCV RNA testing.

Reproducibility. Intrapanel reproducibility could be evaluated by comparison of the results for samples 5 and 8 of panel 2, both containing target levels of 5,000 copies of HCV genotype 1 per ml. Table 2 shows that the percentages of qualitative

TABLE 6. Interpanel reproducibility

| Sample target concn (copies/ml) | Genotype | Parameter | Panel 1 | Panel 2 |
|------------------------------------|----------|---|-------------|-------------|
| 500,000 | 1 | Detection rate, qualitative tests (%) | 97.8 | 100 |
| | | Detection rate, quantitative tests (%) | 100 | 100 |
| | | GM (log ₁₀) ± SD for all data sets | 5.87 ± 0.43 | 5.59 ± 0.34 |
| 50,000 | 3 | % positive tests within acceptable range ^a | 63 | 83 |
| | | Detection rate, qualitative tests (%) | 95.6 | 98.7 |
| | | Detection rate, quantitative tests (%) | 94.3 | 95.8 |
| 5,000 | 1 | GM (log ₁₀) ± SD for all data sets | 5.08 ± 0.55 | 4.97 ± 0.55 |
| | | % positive tests within acceptable range | 64 | 72 |
| | | Detection rate, qualitative tests (%) | 91.1 | 94.7 |
| 200 | 1 | Detection rate, quantitative tests (%) | 80.0 | 77.1 |
| | | GM (log ₁₀) ± SD for all data sets | 3.92 ± 0.51 | 3.93 ± 0.50 |
| | | % positive tests within acceptable range | 82 | 84 |
| 200 | 1 | Detection rate, qualitative tests (%) | 73.3 | 72.0 |
| | | Detection rate, quantitative tests (%) | 34.3 | 10.4 |
| | | GM (log ₁₀) ± SD for all data sets | 3.02 ± 0.80 | 2.89 ± 0.42 |
| | | % positive tests within acceptable range | 75 | 80 |

^a Percentage of positive results within the range GM (log₁₀) ± 0.5.

correct results for samples 5 and 8 were 95% and 92%, respectively, for the true qualitative methods and 77% and 73%, respectively, for the quantitative assays. Table 4 shows that the GMs of the quantitative results obtained for samples 5 and 8 were $3.93 \pm 0.50 \log_{10}$ and $3.78 \pm 0.40 \log_{10}$, with 84% and 89% of the data in the range $GM \pm 0.50 \log_{10}$, respectively.

Interpanel reproducibility could be evaluated from the results obtained with four positive samples included in both panels (Table 6). The results obtained with these samples on both occasions were similar, although the weak-positive sample (target level of 200 copies/ml) was detected by fewer laboratories in the panel 1 group (10.4%) than in the panel 2 group (34.3%) ($P = 0.008$). The percentages of quantitative results within the range of $\pm 0.5 \log_{10}$ of the GM were slightly higher in panel 2 for all four samples.

Comparison of laboratory performances on panel 1 and panel 2. Sixty laboratories submitted either qualitative data sets ($n = 33$) or quantitative data sets ($n = 27$) for both panels. Their scores on both panels were compared. Poor qualitative scores were obtained by 5 of 33 (15.2%) and 0 of 33 (0%) of the qualitative data sets returned for panel 1 and panel 2, respectively (data not shown). Two laboratories, both changing from in-house methods with panel 1 to Roche methods with panel 2, improved their qualitative score enormously, from -1 to 8 and from 3 to 8. In contrast, the percentages of poor qualitative scores obtained with the quantitative data sets increased from 11% (3 of 27) with panel 1 to 25.9% (7 of 27) with panel 2. Finally, the percentages of participants with poor quantitative scores increased slightly from 25.9% (7 of 27) with panel 1 to 33.3% (9 of 27) with panel 2.

DISCUSSION

So far only a few proficiency studies on HCV RNA detection have been published. In an early study, performed in conjunction with the European Expert Group on Viral Hepatitis, 10 plasma samples (four positive and six negative) were qualitatively tested for HCV RNA by 31 laboratories (19). Of the 31 reported data sets, no less than 9 (29%) showed false-positive results and 12 (39%) gave false-negative results. Only 10 (32%) of the data sets had all samples correct. A similar study performed 3 years later with 134 data sets on four positive and six negative samples showed only little improvement: 28 (21%) data sets with false-positive results, but 84 (63%) with false-negative results (1).

The same approach was used for an early proficiency study on the detection of HBV DNA, where 39 laboratories submitted 43 data sets on 12 plasma samples (seven positive and five negative). In that study, 15 (35%) data sets showed false-positive results, 16 (37%) showed false-negative results, and 12 (28%) showed correct results for all samples (9). Similar problems with sensitivity and specificity were reported some years ago for the detection of *Mycobacterium tuberculosis* DNA (8). All these studies clearly demonstrated large numbers of laboratories with sensitivity and specificity problems. In none of these studies was nucleic acid quantitated.

Compared with the studies mentioned above, the present study on HCV RNA shows a much better specificity: only 2.5% of the 80 data sets for panel 1 and 1.6% of the 123 data sets for panel 2 showed false-positive results. Similar low false-positive

rates have recently been found in the EU QCCA proficiency studies on HBV DNA (15), human immunodeficiency virus RNA (A. M. van Loon, J. Schirm, E. Valentine-Thon, J. Reid, P. E. Klapper, and G. M. Cleator, unpublished data), enterovirus RNA (16), and *Chlamydia trachomatis* DNA (R. P. Verkooyen, G. T. Noordhoek, P. E. Klapper, J. Reid, J. Schirm, G. M. Cleator, and G. Hoddevik, unpublished data). In contrast, a recent EU QCCA proficiency study on the detection of herpes simplex virus DNA (L. Schloss, P. Cinque, G. M. Cleator, J.-E. Echevarria, K. I. Falk, P. E. Klapper, J. Schirm, B. F. Vestergaard, H. G. M. Niesters, T. Popow-Kraupp, W. G. V. Quint, A. M. van Loon, and A. Linde, unpublished data) still showed relatively large numbers of data sets with false-positive results (8% in 1999 and 18% in 2000). This may have been related to high viral loads in some of the test samples, which may have caused contamination in the laboratories of some participants, and the fact that all herpes simplex virus DNA testing was still performed by in-house methods. The low false-positivity rates in the other recent studies probably reflect the greater expertise of the participating laboratories in addressing the contamination issue compared to several years ago, coupled with the availability of commercial kits.

For the identification of HCV-infected patients, qualitative HCV RNA detection methods which are as sensitive as possible should be used. For quantitative HCV RNA detection, used for monitoring the efficacy of antiviral therapy, precision and reproducibility are considered most important. Nevertheless, there is also an increasing demand for more sensitive quantitative HCV RNA tests. Unfortunately, the rates of false-negative results are difficult to determine and to compare between studies because of the large variation in the detection limits of the various assays used and the different viral loads of the samples used in the studies. Of the 80 data sets for panel 1, no less than 46% showed negative results on the (low) positive samples (24% of the true qualitative data sets and 73% of the quantitative data sets). Of the 123 data sets for panel 2, the percentage of negative results on (low) positive samples was 58% (35% of the true qualitative data sets and 83% of the quantitative data sets).

This increase in the negativity rate on (low) positive samples and subsequent decrease in overall performance were most probably related to the lower average viral loads in panel 2 combined with the relatively high detection limits of some of the quantitative assays used. The bDNA version 2.0 assay, for example, has a lower detection limit of no less than 200,000 copies/ml, so that only the highest positive sample (target load of 500,000 copies/ml) should be positive. However, it is imaginable that in addition, due to run-to-run variation of the lower detection limit and the lack of standardization (see below), samples with target loads of 50,000 copies/ml can still be detected in the bDNA test. Indeed, 6 of 8 and 9 of 11 of the bDNA data sets on panel 1 and panel 2, respectively, were able to detect all the samples with target levels of 50,000 copies/ml. In addition, even in the samples with target levels of 5,000 copies/ml, HCV RNA was detected by 2 of 8 and 1 of 33 tests performed by users of the bDNA test. However, considering the high lower detection limit of the bDNA version 2.0 test, these three positive results might perhaps be considered false positives.

In the present study, no penalty points were given when

positive samples with $\leq 5,000$ copies/ml were reported negative. Therefore, a result of $< 200,000$ copies/ml for a sample containing 5,000 copies/ml was not penalized, although in our opinion a lower detection limit of 200,000 copies/ml will not meet all the critical requirements of a modern diagnostic laboratory. Fortunately, this opinion is now also shared by industry, since most of the recently developed commercial test systems, including the HCV RNA bDNA version 3.0 test, have lower detection limits of $\leq 1,000$ copies/ml.

The primary purpose of proficiency testing is to determine whether a laboratory is capable of providing reliable results, not whether it is able to carry out a particular (commercial) test adequately. Although, in principle, the type of method used is less relevant, results of proficiency testing may also yield useful information on the performance of the particular methods used. However, it should be noted that there are some important restrictions on the interpretation of such comparisons. This is due to the composition of the panel, the relatively small number of samples, and the small number of laboratories using some of the assays. The interpretation of our quantitative data was inevitably biased because the vast majority of the data sets were obtained with only one test system, the Roche (Cobas)Amplificor Monitor assays. Nevertheless, we consider our quantitative scoring system the most useful and practical approach at the moment, since in practice the most widely used assay will more or less serve as the de facto standard. It does underline, however, the strong need for standardization of all quantitative HCV RNA detection methods, for instance, with the recently introduced World Health Organization standard (2, 11). Interestingly, this World Health Organization standard contains HCV genotype 1 only. It remains to be seen whether standard preparations for other HCV genotypes will also be necessary.

In the present study, the bDNA version 2.0 method gave consistently higher quantitative results than the (Cobas)Amplificor Monitor assays. This is in accordance with earlier data (14) and can be explained by the lack of standardization (11). It was also reported that whereas the Roche Monitor version 2.0 and the bDNA version 2.0 tests showed equal sensitivity for the most common HCV genotypes (7), HCV genotypes other than genotype 1 had been underestimated by earlier versions of the Roche tests (7, 14). This may explain why, in our panel 1, the largest difference between the bDNA test and the Roche tests was found for the sample containing HCV genotype 3.

Notwithstanding the lack of standardization, the intrapanel and interpanel reproducibility of our samples was excellent, except for the detection rate of the weak-positive sample, which was much higher on panel 1 than on panel 2 (Table 6). We cannot explain this difference. Our results also showed that the CV of the results obtained with the bDNA method were significantly smaller than those with the other amplification methods. This is in concordance with studies on human immunodeficiency virus RNA, indicating that signal amplification methods are less susceptible to variation than target amplification methods (13, 17; A. M. van Loon, J. Schirm, E. Valentine-Thon, J. Reid, P. E. Klapper, and G. M. Cleator, unpublished data).

Our study involved large and increasing numbers of participants (57 and 81 in panels 1 and 2, respectively) and data sets (80 and 123, respectively). The percentages of data sets with

good or adequate qualitative scores decreased slightly from 84% in panel 1 to 80% in panel 2, which was probably due to the average lower viral loads in panel 2. In contrast, the percentages of data sets with good or adequate quantitative scores increased from 60% to 73%. This may be due to the more extensive use of later versions of commercial test systems. Laboratories using in-house methods often had poor results, and laboratories changing from in-house methods to commercial methods improved their performance.

In conclusion, our study indicates that considerable improvement of the molecular detection of HCV RNA has been achieved in recent years, particularly through the use of commercial assays. However, further standardization is still needed, and most laboratories should use more sensitive quantitative assays. Finally, the present study underlines that proficiency panels are important tools for monitoring the quality of diagnostic laboratory tests.

ACKNOWLEDGMENTS

The EU Concerted Action Programme was supported with a grant from the EU Biomed 2 Programme.

We thank all participating laboratories and reference laboratories (Roche Diagnostics and Bayer Diagnostics) for their contribution to the HCV RNA proficiency study. We thank Dirk. S. Luijt for helping with evaluation of the data and Kees van Slochteren for statistical analyses.

REFERENCES

- 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.
- Jorgensen, P. A., and P. D. Neuwald. 2001. Standardized hepatitis C virus RNA panels for nucleic acid testing assays. *J. Clin. Virol.* **20**:35-40.
- Kawai, S., O. Yokusaka, T. Kanda, F. Imazeki, Y. Maru, and H. Saisho. 1999. Quantification of hepatitis C virus by TaqMan PCR: comparison with HCV Amplificor monitor assay. *J. Med. Virol.* **58**:121-126.
- Lunel, F., M. Mariotti, P. Cresta, I. De La Croix, J.-M. Huraux, and J.-J. Lefrère. 1995. Comparative study of conventional and novel strategies for the detection of hepatitis C virus RNA in serum: Amplificor, branched-DNA, NASBA and in-house PCR. *J. Virol. Methods* **54**:159-171.
- Martinot-Peignoux, M., N. Boyer, M. Pouteau, C. Castelnau, N. Guily, V. Duchatelle, A. Aupérin, C. Degott, J. D. Benhamou, S. Erlinger, and P. Marcellin. 1998. Predictors of sustained response to alpha interferon therapy in chronic hepatitis C. *J. Hepatol.* **29**:214-223.
- Martinot-Peignoux, M., P. Marcellin, M. Pouteau, C. Castelnau, N. Boyer, M. Poliquen, C. Degott, V. Descombes, V. Lebreton, V. Milotova, J. D. Benhamou, and S. Erlinger. 1995. Pretreatment serum HCV RNA levels and HCV genotype are the main and independent prognostic factors of sustained response to alpha interferon therapy in chronic hepatitis C. *Hepatology* **22**:1050-1056.
- Mellor, J., A. Hawkins, and P. Simmonds. 1999. Genotype dependence of hepatitis C virus load measurement in commercially available quantitative assays. *J. Clin. Microbiol.* **37**:2525-2532.
- 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.
- Quint, W. G. V., R. A. Heijntink, J. Schirm, W. H. Gerlich, and B. Niesters. 1995. Reliability of methods for hepatitis B virus DNA detection. *J. Clin. Microbiol.* **33**:225-228.
- Roth, W. K., J. H. Lee, B. Ruster, and S. Zeuzem. 1996. Comparison of two quantitative hepatitis C virus reverse transcriptase PCR assays. *J. Clin. Microbiol.* **34**:261-264.
- 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.
- Schröter, M., B. Zöllner, P. Schäfer, R. Laufs, and H.-H. Feucht. 2001. Quantitative detection of hepatitis C virus RNA by light cycle PCR and comparison with two different PCR assays. *J. Clin. Microbiol.* **39**:765-768.
- Schuurman, R., D. Descamps, G. J. Weverling, S. Kaye, J. Tijnagel, I. Williams, R. van Leeuwen, R. Tedder, C. A. B. Boucher, F. Brun-Vezinet, and

- C. Loveday. 1996. Multicenter comparison of three commercial methods for quantitation of human immunodeficiency virus type 1 RNA in plasma. *J. Clin. Microbiol.* **34**:3016–3022.
14. Tong, C. Y. W., R. C. Hollingsworth, H. Williams, W. L. Irving, and I. T. Gilmore. 1998. Effect of genotypes on quantification of hepatitis C virus (HCV) RNA in clinical samples with the Amplicor HCV Monitor test and the Quantiplex HCV RNA 2.0 assay (bDNA). *J. Med. Virol.* **55**:191–196.
15. Valentine-Thon E., A. M. van Loon, J. Schirm, J. Reid, P. E. Klapper, and G. M. Cleator. 2001. A European proficiency testing program for molecular detection and quantitation of hepatitis B virus DNA. *J. Clin. Microbiol.* **39**:4407–4412.
16. 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.
17. Yen-Lieberman, B., D. Brambilla, B. Jackson, J. Bremer, R. Coombs, M. Cronin, S. Herman, D. Katzenstein, S. Leung, H. J. Lin, P. Palumbo, S. Rasheed, J. Todd, M. Vahey, and P. Reichelderfer. 1996. Evaluation of a quality assurance program for the quantitation of human immunodeficiency virus type 1 RNA in plasma by the AIDS Clinical Trials Group Virological Laboratories. *J. Clin. Microbiol.* **34**:2695–2701.
18. Yu, M.-L., W.-L. Chuang, C.-Y. Dai, S.-C. Chen, Z.-Y. Lin, M.-Y. Hsieh, L.-Y. Wang, and W.-Y. Chang. 2000. Clinical evaluation of the automated Cobas Amplicor HCV monitor test version 2.0 for quantifying serum hepatitis C virus RNA and comparison to the quantiplex HCV version 2.0 test. *J. Clin. Microbiol.* **38**:2933–2939.
19. 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.