

Reliability of Nucleic Acid Amplification Methods for Detection of *Chlamydia trachomatis* in Urine: Results of the First International Collaborative Quality Control Study among 96 Laboratories

Roel P. Verkooyen,¹ Gerda T. Noordhoek,^{2*} Paul E. Klapper,³ Jim Reid,⁴ Jurjen Schirm,⁵
Graham M. Cleator,⁶ Margareta Ieven,⁷ and Gunnar Hoddevik⁸

Department of Medical Microbiology and Infectious Diseases, Erasmus MC, Rotterdam,¹ Public Health Laboratory Friesland, Leeuwarden,² and Regional Public Health Laboratory, Groningen,⁵ The Netherlands; Public Health Laboratory, Leeds,³ ChimaeraBio, Glasgow,⁴ and University of Manchester, Manchester,⁶ United Kingdom; Department of Microbiology, University of Antwerp UIA, Antwerp, Belgium⁷; and National Institute of Public Health, Oslo, Norway⁸

Received 22 July 2002/Returned for modification 2 October 2002/Accepted 4 April 2003

The first European Quality Control Concerted Action study was organized to assess the ability of laboratories to detect *Chlamydia trachomatis* in a panel of urine samples by nucleic acid amplification tests (NATs). The panel consisted of lyophilized urine samples, including three negative, two strongly positive, and five weakly positive samples. Ninety-six laboratories in 22 countries participated with a total of 102 data sets. Of 204 strongly positive samples 199 (97.5%) were correctly reported, and of 506 weakly positive samples 466 (92.1%) were correctly reported. In 74 (72.5%) data sets correct results were reported on all samples, and 17 data sets (16.7%) showed either one false-negative or one false-positive result. In another 11 data sets, two or more incorrect results were reported, and two data sets reported a false-positive result on one negative sample. The Roche COBAS Amplicor test was performed in 44 (43%) data sets, the Abbott LCx assay was performed in 31 (30%) data sets, the Roche Amplicor manual assay was performed in 9 (9%) data sets, an in-house PCR was performed in 9 (9%) data sets, the Becton Dickinson ProbeTec ET assay was performed in 5 (4.9%) data sets, and the GenProbe TMA assay was performed in 4 (3.9%) data sets. The results of the Roche Amplicor manual (95.6% correct), COBAS Amplicor (97.0%), and Abbott LCx (94.8%) tests were comparable ($P = 0.48$). The results with the in-house PCR, BD ProbeTec ET, and GenProbe TMA tests were reported correctly in 88.6, 98, and 92.5% of the tests, respectively. Freeze-drying of clinical urine specimens proved to be a successful method for generating standardized, stable, and easy-to-transport samples for the detection of *C. trachomatis* by using NATs. Although the results, especially the specificity, for this proficiency panel were better than most quality control studies, sensitivity problems occurred frequently, underlining the need for good laboratory practice and reference reagents to monitor the performance of these assays.

A multiplicity of nucleic acid amplification test (NAT) procedures are now available for the diagnosis of *Chlamydia trachomatis* infections. *C. trachomatis* amplification tests can be performed on endocervical or urethral swab specimens and urines (6). In addition to in-house PCR tests, commercial NAT assays are available, such as Abbott LCx, Becton Dickinson ProbeTec ET, GenProbe TMA, and Roche Amplicor (COBAS and manual) PCRs. All tests aim to be rapid, sensitive, specific, and easy to perform. However, previous studies have shown that application of NATs may be unreliable because of cross-contamination, inappropriate treatment of the clinical samples leading to the loss of target DNA, or the presence of inhibitors in the sample (1, 3, 8, 11).

Members of the European Union Concerted Action on Quality Control (EU-QCCA) of Nucleic Acid Amplification in Diagnostic Virology and members of the Study Group on Molecular Diagnostics of the European Society for Clinical Microbiology and Infectious Diseases formed a working party to establish a quality assessment program for the evaluation of

currently used NATs for the detection of *C. trachomatis*. The aim of the study was to assess the quality performance of laboratories for detection of *C. trachomatis* in a panel of samples that could be tested with all existing commercial and in-house test systems and that would resemble clinical samples. Since commercial systems require their specific transport medium for swab specimens, urine samples were chosen as the most suitable specimen for the EU-QCCA panel. All commercial systems include a protocol for testing *C. trachomatis* in urine samples.

In order to avoid the high costs associated with shipment of infectious goods on dry ice, lyophilized urine specimens were prepared.

MATERIALS AND METHODS

Collection of urine samples. Negative urine samples were obtained by collecting up to 50 ml of early morning first-catch urine samples over a period of 3 months from two healthy females (urine samples B and C) and one healthy male volunteer (urine sample D). Urine samples from each volunteer were pooled. The volunteers had no history of infection with *C. trachomatis* or *N. gonorrhoeae*, and the pools were examined for these microorganisms by culture and NAT. These tests were all negative. A mixture of urine types was formed by mixing equal amounts of the pooled urine samples (mix A).

First-catch urine samples from patients visiting the clinics for sexually transmitted diseases of the Erasmus Medical Center in Rotterdam, The Netherlands, and of the National Institute of Public Health in Oslo, Norway, were collected

* Corresponding author. Mailing address: Public Health Laboratory Friesland, Post Box 21020, 8900 JA Leeuwarden, The Netherlands. Phone: 31-58-2939495. Fax: 31-58-2939200. E-mail: gerda.noordhoek@lvf.nl.

TABLE 1. Composition of the panel and overview of correct results per sample

Sample no.	Composition of the sample		Expected result	No. of samples tested	No. of correct results (%)
	Background	Positive urine added ^b			
2	Urine B		Negative	101	100 (99.0)
5	Urine C		Negative	102	102 (100)
9	Urine D		Negative	101	100 (99.0)
1	Mix A ^a	I	Strongly positive	102	100 (98.0)
4	Mix A	II	Strongly positive	102	99 (97.1)
3	Urine B	I	Weakly positive	101	95 (94.1)
6	Urine C	I	Weakly positive	102	92 (90.2)
7	Urine C	I	Weakly positive	101	93 (92.1)
8	Mix A	I	Weakly positive	102	97 (95.1)
10	Mix A	II	Weakly positive	100	89 (89.0)

^a Mix A is a mixture of equal volumes of urine samples B, C, and D.

^b Positive urine pool I contained *C. trachomatis* and *N. gonorrhoeae*. Pool II contained *C. trachomatis*.

and stored at 4°C. Samples were tested for the presence *C. trachomatis* DNA and *Neisseria gonorrhoeae* DNA with the Roche COBAS Amplicor PCR and the Abbott LCx tests according to the manufacturer's procedures. *C. trachomatis*-positive urine samples were diluted 10⁻¹ and 10⁻² and retested the same day. Urine samples positive at the 10⁻² dilution were considered strongly positive. After estimation of the volumes, all urine samples were frozen at -70°C, with two extra aliquots of 1 ml for further pilot testing. It appeared that some of the urine samples from Rotterdam, but not those from Oslo, contained *N. gonorrhoeae* DNA in addition to *C. trachomatis* DNA.

All positive urine samples were thawed, and two different pools were prepared. The 28 urine samples collected in Rotterdam were used to prepare a *C. trachomatis*-positive-*N. gonorrhoeae*-positive urine pool (pool I). The 22 urine samples collected in Oslo were used to prepare a *C. trachomatis*-positive-*N. gonorrhoeae*-negative urine pool (pool II). Tenfold dilutions of pools I and II were prepared in the *C. trachomatis*-*N. gonorrhoeae*-negative urine mix, and 1.7-ml aliquots were freeze-dried by using a Leybold-Hereaus GT 2 freeze-drier and housed in a temperature-controlled environment (17°C).

Quality control assurance and production of the panel. The lowest dose 50% endpoint (LD50) was defined as the lowest concentration of target that gave a 50% chance of a positive test result. The 10-fold dilutions of urine pools I and II were tested six times by each of four different commercial test systems: Abbott LCx, BD ProbeTec ET, COBAS Amplicor PCR, and GenProbe TMA. Three ampoule contents were pooled to provide the 4-ml volume required by the BD ProbeTec ET test. The LD50 value of a positive response in each test was calculated by the method of Kärber (7). The weakly positive samples in the final panel contained 10 times the concentration of the LD50.

The final panel consisted of 10 1.7-ml portions of lyophilized urine samples: 3 negative samples, 2 strongly positive samples, and 5 weakly positive samples. Three different urine samples from healthy volunteers were used for dilution of the positive pool I and II to generate a heterogeneous panel (Table 1). The samples were coded and randomized.

Quality assurance for the quality control panel was performed by four reference laboratories, and tests were performed with Abbott LCx, BD ProbeTec ET, COBAS Amplicor PCR, and GenProbe TMA. Laboratories performing the BD ProbeTec assay received three vials of each specimen; these had to be pooled after restoring the volume to 1.7 ml to obtain the requested specimen volume of 4 ml.

Organization of the study. The proficiency panel was distributed from the production laboratory in Oslo, Norway, by surface mail at ambient temperature to the participants. The packet also included an information sheet with instructions as to how to open the vials and restore the volume to 1.7 ml. Reporting of arrival, results, questionnaire, coding, and confidentiality was organized as described previously (15). The Neutral Office in Manchester, United Kingdom, received all results. Data were analyzed anonymously in Leeuwarden and Rotterdam, The Netherlands, by using the Statistical Package for the Social Science (SPSS version 10). Statistical evaluation of the collected data was performed by using the Fisher exact test; probability tests were two-tailed.

RESULTS

Validation of freeze-drying of *C. trachomatis*-infected urine specimens. A set of 1-ml portions of 10-fold dilutions of a pool

of strongly positive urine samples was lyophilized, randomized, and sent to other laboratories to test for the presence of *C. trachomatis* DNA or RNA with COBAS Amplicor PCR (five laboratories), Abbott LCx (four laboratories), GeneProbe TMA (two laboratories), and BD ProbeTec ET (one laboratory); one set of samples was tested with the Vidas Probe assay (Biomerieux, Boston, Mass.). Reproducible results were received from duplicate series tested with the same method at the same or at different laboratories. The sensitivity of the COBAS Amplicor, LCx, Vidas Probe, and BD ProbeTec ET assays were similar: dilutions 10⁻³ to 10⁻⁴ tested positive. The laboratories performing the GenProbe TMA assay reported a lower sensitivity (10⁻² to 10⁻³); this probably resulted from the use of a 1-ml specimen, whereas the provider prescribes 1.5-ml of urine to be tested. Samples from the same batch were stored at room temperature and retested 5 months and 1 year later with COBAS Amplicor and LCx. Comparable results were obtained that confirmed that the lyophilized samples were stable for at least 12 months.

Distribution and response. The panel was sent to 105 laboratories in 22 countries. The median transport time was 3 days, with a range of 1 to 36 days. Results were received from 96 participants (91.4%). Ninety laboratories returned one data set, and six participants performed two tests and returned two data sets, which yielded a total of 102 evaluable data sets.

Performance of the laboratories. In total, 304 negative samples were tested. The composition of the panel and the results on the individual samples are presented in Table 1. Only two laboratories reported one false-positive result; these were both experienced laboratories examining more than 2,000 samples per year. Of the 204 strongly positive samples, 199 (97.5%) were reported correctly; of 506 examinations of weakly positive samples, 466 (92.1%) were reported correctly.

In 74 (72.5%) data sets all samples were reported correctly. In 17 (16.7%) data sets one incorrect result was reported, in 16 of these 1 weakly positive sample was reported as a false-negative result. In six (5.9%) data sets two weakly positive samples were reported incorrectly; one (1%) data set missed three weakly positive samples. In three (2.9%) data sets two or three weakly positive samples plus one or two strongly positive samples were reported incorrectly, and one (1%) data set in-

TABLE 2. Comparison of results and type of assay

Method	No. of data sets	No. of samples with correct results/no. of samples tested (%)			
		Negative samples	Strongly positive samples	Weakly positive samples	All samples
COBAS Amplicor	44	131/132 (99.2)	87/88 (98.9)	207/218 (95.0)	425/438 (97.0)
Abbott LCx	31	92/93 (98.9)	59/62 (95.2)	142/154 (92.2)	293/309 (94.8)
Amplicor manual	9	27/27 (100.0)	18/18 (100.0)	41/45 (91.1)	86/90 (95.6)
PCR (in house)	9	25/25 (100.0)	17/18 (94.4)	36/45 (80.0)	78/88 (88.6)
BD Probe Tec ET	5	15/15 (100.0)	10/10 (100.0)	23/24 (95.8)	48/49 (98.0)
GenProbe TMA	4	12/12 (100.0)	8/8 (100.0)	17/20 (85.0)	37/40 (92.5)
Total	102	302/304 (99.3)	199/204 (97.5)	466/506 (92.1)	967/1014 (95.4)

correctly reported one weakly positive sample, one strongly positive sample, and one negative sample.

Relation between score, type of assay, and procedures. Table 2 presents an overview of the scores classified by the type of assay used. The BD ProbeTec ET has the highest rate (98.0%) of correct results; however, only five participants used this new assay. Equal percentages of correct results were reported with the two Roche tests—Amplicor manual (95.6%) and COBAS Amplicor (97.0%)—and with the Abbott LCx test (94.8%) ($P = 0.48$). The GenProbe TMA test scored lower (92.5%) because only 85% of the results of the weakly positive samples were correctly reported. This result was in accordance with the expected performance of the test as predicted by the random chance calculation of the results from pilot tests and LD50 calculations (data not shown). The in-house PCR tests had the lowest score (88.6% correct results).

Application of an internal control to monitor inhibition of the reaction in each sample was reported by 42 data sets. Three participants, two using COBAS-Amplicor and one using in-house PCR, reported one or more samples with inhibition. When these samples were tested again, the inhibition problem was solved in the COBAS Amplicor assay but not in the in-house PCR assay.

The majority of participants (91.7%) were laboratories that carry out NAT for clinical diagnostics, 76 (79.2%) were hospital or public health laboratories, and 12 (12.5%) were commercial (private) laboratories. Five (5.2%) laboratories mentioned that they were manufacturers of NAT assays. Three participants did not answer this question. Commercial (private) laboratories tended to score lower than diagnostic laboratories and manufacturers (92.8, 95.7, and 100%, respectively; $P = 0.08$).

The number of NATs performed for detection of *C. trachomatis* on an annual basis varied from one laboratory performing 10 to 100 tests up to 21 laboratories examining more than 10,000 samples. More than 80% of the respondents were experienced users of NATs (>1,000 tests per year). However, there was no correlation between the number of tests performed annually and the quality of the results.

Four participants performed NAT in a single laboratory room; data sets from these labs showed a higher number of incorrect results, i.e., 6 of 39 (15.4%) versus 41 of 945 (4.3%) for those using more than one room ($P = 0.008$).

DISCUSSION

This is the first international quality control study for the detection of *C. trachomatis* based on molecular diagnostic

methods and lyophilized urine samples. The pilot studies and the results from the final panel show that freeze-drying of clinical urine samples for NATs is a successful method for obtaining stable, standardized samples that are easy to transport.

The panel was composed to resemble clinical samples, with strongly and weakly positive urine samples from women and men. The weakly positive samples contained approximately the same number of *C. trachomatis* organisms believed to be present in asymptomatic infected humans (R. P. Verkooyen, unpublished data). Since more than 60% of all female *C. trachomatis* infections are asymptomatic, the working party thought it important that participants should be able to detect these low numbers of bacteria. Also, population-based screening for *C. trachomatis* is expected to detect many asymptomatic infections (2, 5, 13).

Since double infection with *C. trachomatis* and *N. gonorrhoeae* has been observed frequently (9, 14), we investigated whether the tests were able to cope with samples containing both *C. trachomatis* and *N. gonorrhoeae*. The results showed that the presence of these dual targets did not adversely affect test sensitivity for *C. trachomatis* (Table 1).

The scores described here were very good; 89.2% of data sets scored 90% or higher. However, five (5%) data sets scored $\leq 70\%$. The primary purpose of the proficiency testing was to estimate whether a laboratory is capable of providing reliable results and not to test the performance of different commercial assays. However, as the number of participants applying certain commercial assays was high, the results also yielded useful information on the assays used. It showed that there was no significant difference between the Roche PCR and the Abbott LCx tests, although the mean score from the LCx test was 2.2% lower. This difference was mainly due to the errors with the weakly positive samples. Results from the pilot tests suggest that COBAS Amplicor and LCx tests have the same chance of correctly reporting weakly positive samples. The fact that the results from participants in study performing the LCx assay were slightly less successful than with the COBAS assay may be due to inhibition of the amplification reaction sample (1, 3, 8, 11). From previous investigations we learned that the accuracy of molecular tests with urine samples may be compromised if urine residues remain after centrifugation; such a problem would not be detected in the LCx system since this system lacks an internal inhibition control. On the other hand, a weakly positive sample may also be missed when the pellet is lost after

centrifugation. In that case, there is no inhibition and a false-negative test result is reported.

Previous quality control studies revealed that the number of manually performed pretreatment steps and the difficulty of processing samples before samples enter an automatic system are the major causes of false-positive and false-negative results (4, 10, 12, 17). The procedures used for the detection of *C. trachomatis* in urine samples have relatively simple pretreatment protocols, which may explain the differences in results between this quality control study and the other studies. The present study shows a very low number of false-positive results, and only two negative samples were reported as positive in two different data sets. These low rates were also reported in recent quality control studies and may reflect the greater expertise of the participating laboratories compared to several years ago and the increasing use of commercial kits in general (15, 16).

In the present study, weakly positive samples were missed approximately three times more frequently than the strongly positive samples, indicating a problem in test sensitivity. The question always arises as to whether in-house PCR amplification is suitable for routine clinical microbiological diagnostics. The group using in-house PCR tests had the lowest mean score (88.6% correct results), which shows that results obtained by in-house amplification procedures should be interpreted with care.

For each of the commercial methods the manufacturer prescribes the volume of sample to be used. Participants were asked to calculate the sample volume equivalent used in the amplification reaction and to report whether they adhered to the manufacturer's protocol. These answers were not always in concordance; some participants calculated a different volume than the one prescribed but answered that they used the manufacturer's protocol. In some cases, only 50% of the prescribed volumes for reagents and sample were used. Use of a smaller or a larger volume may compromise the sensitivity of the method or increase the inhibition rate. Ten participants apparently used lower equivalents of urine in the test than are prescribed by the manufacturer, resulting in a significant decrease in sensitivity, compared to the participants who used the prescribed sample equivalents ($P = 0.036$). Only three participants violated the test protocol by increasing the equivalent sample volume used in the test. No significant difference was observed in these cases.

Participation from different European countries was low in comparison to the number of laboratories performing NATs for the detection of *C. trachomatis*. Some laboratories may underestimate the importance of external quality assessment as part of quality management systems. At the time of the present study, only 31.2% of participant laboratories were accredited. We expect that the need for QC panels will increase if in more countries accreditation becomes obligatory. The outcome of the present study underlines the need for reference reagents and standardized operating procedures, which would enable experienced technicians to perform reliable quality control assessment of nucleic acid amplification methods. The first steps are taken toward a standardized procedure for quality control of nucleic acid amplification assays for the detection of *C. trachomatis*, which should be available for every laboratory performing NATs in routine clinical practice.

ACKNOWLEDGMENTS

We thank Abbott Diagnostics, Roche Diagnostics, Biomerieux, and Becton Dickinson for support. The EU-QCCA Program was supported by a grant from the EU Biomed 2 Program.

We are grateful to J. Mouton (Canisius Hospital, Nijmegen, The Netherlands), P. Oostvogel (Westeinde Hospital, The Hague, The Netherlands), A. van den Brule (Free University of Amsterdam, Amsterdam, The Netherlands), J. Thijie (Academic Hospital Maastricht, Maastricht, The Netherlands), C. Roemer (Medical Microbiological Laboratory, Utrecht, The Netherlands), L. Sabbe (Public Health Laboratory Zeeland, Goes, The Netherlands), and M. de Vries (LCPL, Leiden, The Netherlands) for performance of pilot experiments and reference testing. We also thank Margaret Barringer for all administrative work and all participating laboratories.

REFERENCES

- Berg, E. S., G. Anestad, H. Moi, G. Storvold, and K. Skaug. 1997. False-negative results of a ligase chain reaction assay to detect *Chlamydia trachomatis* due to inhibitors in urine. *Eur. J. Clin. Microbiol. Infect. Dis.* **16**:727-731.
- Brady, M., C. Baker, and L. S. Neinstein. 1988. Asymptomatic *Chlamydia trachomatis* infections in teenage males. *J. Adolesc. Health Care* **9**:72-75.
- Chernesky, M. A., D. Jang, J. Sellors, K. Luinstra, S. Chong, S. Castriciano, and J. B. Mahony. 1997. Urinary inhibitors of polymerase chain reaction and ligase chain reaction and testing of multiple specimens may contribute to lower assay sensitivities for diagnosing *Chlamydia trachomatis* infected women. *Mol. Cell. Probes* **11**:243-249.
- Damen, M., H. T. M. Cuypers, H. L. Zaaier, H. W. Reesink, W. P. Schaasberg, W. H. Gerlich, H. G. M. Niesters, and P. N. Lelie. 1996. International collaborative study on the second EUROHEP HCV-RNA reference panel. *J. Virol. Methods* **58**:175-185.
- Forsey, J. P., E. O. Caul, I. D. Paul, and M. G. Hull. 1990. *Chlamydia trachomatis*, tubal disease and the incidence of symptomatic and asymptomatic infection following hysterosalpingography. *Hum. Reprod.* **5**:444-447.
- Goessens, W. H., J. W. Mouton, W. I. van der Meijden, S. Deelen, T. H. van Rijsoort-Vos, N. Lemmens-den Toom, H. A. Verbrugh, and R. P. Verkooyen. 1997. Comparison of three commercially available amplification assays, AMP CT, LCx, and COBAS Amplicor, for detection of *Chlamydia trachomatis* in first-void urine. *J. Clin. Microbiol.* **35**:2628-2633.
- Kärber, G. 1931. Beitrag zur kollektiven behandlung pharmakologischer reihenversuche. *Arch. Exp. Pathol. Pharmacol.* **162**:480-483.
- Mahony, J., S. Chong, D. Jang, K. Luinstra, M. Faught, D. Dalby, J. Sellors, and M. Chernesky. 1998. Urine specimens from pregnant and nonpregnant women inhibitory to amplification of *Chlamydia trachomatis* nucleic acid by PCR, ligase chain reaction, and transcription-mediated amplification: identification of urinary substances associated with inhibition and removal of inhibitory activity. *J. Clin. Microbiol.* **36**:3122-3126.
- Mardh, P. A., K. Tchoudomirova, S. Elshibly, and D. Hellberg. 1998. Symptoms and signs in single and mixed genital infections. *Int. J. Gynaecol. Obstet.* **63**:145-152.
- Noordhoek, G. T., J. D. A. van Embden, and A. H. J. Kolk. 1996. Reliability of nucleic acid amplification for detection of *Mycobacterium tuberculosis*: an international collaborative quality control study among 30 laboratories. *J. Clin. Microbiol.* **34**:2522-2525.
- Notomi, T., Y. Ikeda, A. Okadome, and A. Nagayama. 1998. The inhibitory effect of phosphate on the ligase chain reaction used for detecting *Chlamydia trachomatis*. *J. Clin. Pathol.* **51**:306-308.
- Quint, W. G. V., R. A. Heijntink, J. Schirm, W. H. Gerlich, and H. G. M. Niesters. 1995. Reliability of methods for Hepatitis B virus DNA detection. *J. Clin. Microbiol.* **33**:225-228.
- Tait, I. A., S. J. Duthie, and D. Taylor-Robinson. 1997. Silent upper genital tract chlamydial infection and disease in women. *Int. J. STD AIDS* **8**:329-331.
- Tapsall, J. W., and M. Kinchington. 1996. The frequency of coinfection with *Neisseria gonorrhoeae* and *Chlamydia trachomatis* in men and women in eastern Sydney. *Pathology* **28**:84-87.
- Valentine-Thon, E., A. M. van Loon, J. Schirm, J. Reid, P. E. Klapper, and G. M. Cleator. 2001. European proficiency testing program for molecular detection and quantitation of hepatitis B virus DNA. *J. Clin. Microbiol.* **39**:4407-4412.
- 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.
- Zaaier, 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.