JOURNAL OF CLINICAL MICROBIOLOGY, Oct. 2007, p. 3184–3190 Vol. 45, No. 10
0095-1137/07/$08.00+0 doi:10.1128/JCM.00789-07

Evolution of an International External Quality Assurance Model To Support Laboratory Investigation of *Streptococcus pneumoniae*, Developed for the SIREVA Project in Latin America, from 1993 to 2005▽

Marguerite Lovgren,[1]* James A. Talbot,[1]† Maria Cristina Brandileone,[2] Silvana T. Casagrande,[2] Clara Inés Agudelo,[3] Elizabeth Castañeda,[3] Mabel Regueira,[4] Alejandra Corso,[4] Ingrid Heitmann,[5] Aurora Maldonado,[5] Gabriela Echániz-Avilés,[6] Araceli Soto-Noguerón,[6] María Hortal,[7] Teresa Camou,[7] Jean-Marc Gabastou,[8] José Luis Di Fabio,[8] and the SIREVA Study Group‡

National Centre for Streptococcus, Edmonton, Alberta, Canada[1]; Instituto Adolfo Lutz, São Paulo, Brazil[2]; Instituto Nacional de Salud, Bogotá, Colombia[3]; Instituto Nacional de Enfermedades Infecciosas ANLIS Carlos Malbrán, Buenos Aires, Argentina[4]; Instituto de Salud Pública, Santiago de Chile, Chile[5]; Instituto Nacional de Salud Pública, Cuernavaca, Mexico[6]; Departamento de Laboratorios de Salud Pública, Montevideo, Uruguay[7]; and Area of Technology and Health Services Delivery, Unit of Essential Medicines, Vaccines and Health Technology, Pan American Health Organization/ World Health Organization, Washington, DC[8]

Received 12 April 2007/Returned for modification 29 May 2007/Accepted 1 August 2007

In 1993 the Pan American Health Organization initiated a laboratory-based surveillance system, called the SIREVA project, to learn about *Streptococcus pneumoniae* invasive disease in Latin American children. In 1994, National Laboratories in six countries were trained to perform serotyping and antibiotic susceptibility testing using broth microdilution to determine the MIC for specified antibiotics. An international External Quality Assurance (EQA) program was developed to monitor and support ongoing laboratory performance. The EQA program was coordinated by the National Centre for Streptococcus (NCS), Edmonton, Canada, and included external proficiency testing (EPT) and a validation process requiring regular submission of a sample of isolates from each laboratory to the NCS for verification of the serotype and MIC. In 1999, the EQA program was decentralized to use three of the original laboratories as regional quality control centers to address operational concerns and to accommodate the growth of the laboratory network to more than 20 countries including the Caribbean region. The overall EPT serotyping accuracies for phase I (1993 to 1998) and phase II (1999 to 2005) were 88.0 and 93.8%, respectively; the MIC correlations within ±1 $\log_2$ dilution of the expected result were 83.0 and 91.0% and the interpretive category agreements were 89.1 and 95.3%. Overall, the validation process serotyping accuracies for phases I and II were 81.9 and 88.1%, respectively, 80.4 and 90.5% for MIC agreement, and 85.8 and 94.3% for category agreement. These results indicate a high level of testing accuracy in participating National Laboratories and a sustained increase in EQA participation in Latin America and the Caribbean.

The impact of childhood pneumonia caused by *Streptococcus pneumoniae* in the developing world is well recognized; however, until recently, the epidemiology of invasive pneumococcal disease in Latin America was poorly documented (3, 4, 8). The growing rates of antibiotic-resistant *S. pneumoniae* worldwide, combined with the promise of new conjugate pneumococcal vaccines, prompted the Pan American Health Organization/World Health Organization (PAHO/WHO) to initiate an international laboratory-based surveillance network that would provide prospective regional serotype prevalence data, antibiotic resistance rates, and epidemiology of invasive pneumococcal disease in Latin American children. In 1993, with significant funding from the Canadian International Development Agency, the National Centre for Streptococcus (NCS), Edmonton, Alberta, Canada, and the Laboratory Centre for Disease Control, Ottawa, Ontario, Canada, were recruited to provide laboratory and epidemiological support for this new project (3, 4, 8). As part of the PAHO/WHO Sistema Regional de Vacunas, the project became known as SIREVA (8).

The establishment of specific goals to guide the SIREVA project was integral to successful development of the quality program: (i) to produce study data accepted as valid by the international community; (ii) to create an intra/inter-regional network based on principles of quality assurance, quality control, and total quality management; (iii) to create a bank of biologically valuable material, including isolates and attached demographic information, to track national epidemiologic trends, including the effects of vaccine intervention; (iv) to use the isolates and demographic information to attract international research interest to strengthen the expertise and knowledge of the regional participants while maintaining their ownership of the isolates; (v) to create and maintain long-term

* Corresponding author. Mailing address: National Centre for Streptococcus, Rm. 1B3.26 WMC, Provincial Public Health Laboratory, 8440-112 St., Edmonton, T6G 2J2 Alberta, Canada. Phone: (780) 407-8977. Fax: (780) 407-8984. E-mail: m.lovgren@provlab.ab.ca.

† Present address: Capital Health, Edmonton, Alberta, Canada.

‡ Contributing members of the study group are listed in the Acknowledgments.

▽ Published ahead of print on 8 August 2007.

relationships between international public health laboratories to share knowledge and expertise; and (vi) to create and maintain relationships within each country between public health laboratories, clinical laboratories, clinicians, and epidemiologists to improve surveillance and quality assurance.

The SIREVA project represented the first international, prospective surveillance program for invasive pneumococcal disease in the developing world (8). It was an ambitious undertaking in view of the limited laboratory infrastructure throughout the region and was made more challenging by our goal to build regional laboratory self-sufficiency rather than to simply extract data or isolates (3). This report describes the development of the external quality assurance (EQA) model through its initial stages, and adjustments that were made based on the experience and growth of the network over time. Phase I describes the original model used over the first 5 years of the project (1994 to 1998); phase II (1999 to 2005) describes the redesigned model that was implemented in 1999 in response to the experience gained during phase I.

MATERIALS AND METHODS

Phase I. National Laboratories in six countries (Argentina, Brazil, Colombia, Chile, Mexico, and Uruguay) were selected based on access to a pediatric population as a source of invasive pneumococcal isolates and the willingness and ability to participate according to the standardized protocol, including participation in a defined EQA program (4, 8).

To ensure consistency in laboratory testing, a workshop was conducted at the NCS in February 1994 to train key personnel from each of the participating countries. Technical instruction and laboratory manuals containing standardized methodologies and relevant quality control testing were provided by the NCS. Pneumococcal identification included optochin sensitivity and bile solubility testing (7). Serotyping was performed by Quellung reaction (10) using commercial antisera (Statens Seruminstitut, Copenhagen, Denmark). Antibiotic susceptibility testing by disk and broth microdilution (BMD) methods was based on Clinical and Laboratory Standards Institute (CLSI; formerly National Committee for Clinical Laboratory Standards [NCCLS]) guidelines that were current at the time of training (11, 12) and updated annually as revised documents were published. Each of the key trainers subsequently held similar regional workshops for their staff and for front line laboratories to optimize culture and isolation techniques in each country.

The EQA program was developed and implemented after completion of the regional training and had two components: an external proficiency testing (EPT) program and a validation process. Since some of the laboratories had no previous experience with EQA, it was important to ensure that the basic factors that were used to guide this process were understood by all participants, namely: (i) the level of challenge appropriate to participating laboratories; (ii) use as an educational tool to deliver new information and to assist with problem solving when difficulties were encountered; (iii) a means of evaluation for "routine" laboratory procedures; (iv) monitoring of the performance of reagents, technical precision and accuracy, and clerical processes (including accurate documentation), as well as identifying staff training opportunities; and (v) the provision of timely feedback to all participants, supporting continuous improvement rather than focusing on "testing failure."

The EPT was coordinated from Canada by the NCS. A quality assurance panel (QAP) consisting of five pneumococcal isolates was distributed once in 1994, three times in 1995, and twice in 1996 to each of the six participating laboratories. Challenge strains were selected to represent a range of serotypes and various antibiotic resistance patterns. The participants were required (i) to verify that the identification as *S. pneumoniae* was correct, (ii) to perform serotyping, (iii) to screen for reduced penicillin susceptibility using a 1-μg oxacillin disk, and (iv) to perform testing to determine the MIC for specified antibiotics. Participants were asked to determine the penicillin MIC for the first distribution. Subsequent distributions requested MICs for cefotaxime and/or ceftriaxone chloramphenicol, erythromycin, trimethoprim-sulfamethoxazole, and vancomycin. Testing of these additional antibiotics was performed at the discretion of the participating laboratory, based on their routine testing algorithms and available resources.

Serotyping was considered to be correct if the group and/or type reported and the NCS result were in agreement. A "group" result was graded as correct if the "factor" sera required for "type" classification were not available in the participating laboratory. Susceptibility results were compared to the NCS result. Oxacillin screening results were evaluated based on correct interpretation as penicillin susceptible (zone size, $\geq$20 mm) or reduced susceptibility to penicillin (zone size, $\leq$19 mm) (12). MIC data were considered correct if the value was within 1 $\log_2$ dilution of the NCS MIC. Interpretive category agreement as susceptible, intermediate, or resistant was evaluated against the category interpretation of the NCS MIC using the NCCLS document that was current at the time of testing. Category interpretation errors were classified as very major, major, and minor consistent with generally accepted definitions (6). Only the antibiotics reported by each participant were evaluated; thus, the denominator used to calculate MIC correlation and category agreement varied by laboratory. The results were summarized, graded, and discussed in a summary report compiled by the NCS and distributed back to each of the participating countries.

The validation process required the submission of a specified sample of clinical pneumococcal isolates to the NCS for verification of the submitting laboratories' serotype and antibiotic susceptibility profiles. Participants were asked to submit all pneumococcal isolates that they were unable to type (this varied according to the inventory of antisera available in each laboratory), 10% of the isolates that were "typeable," and all isolates with reduced penicillin susceptibility as determined by the oxacillin screen for verification of the penicillin MIC. Regular submission of isolates was expected, but a shipping schedule was not defined. Evaluation criteria were consistent with those used for the EPT program.

Phase II. In 1999, the EQA program was decentralized, producing one program, coordinated by the NCS for three quality control centers (QCC), located in Brazil, Colombia, and Mexico, selected after the phase I evaluation. The EQA program for the expanded SIREVA network (over 20 countries) was coordinated locally by the QCC. All countries continued to participate in this decentralized model, but only the EQA program coordinated by the NCS is discussed here.

The original two-part EQA model consisting of EPT and validation process was maintained with minor modifications. The revised EPT program provided two QAP distributions annually to each of the QCC, each containing 10 isolates. Participants were asked to confirm the identification and perform serotyping, oxacillin screening, and MIC testing for penicillin, cefotaxime or ceftriaxone, chloramphenicol, erythromycin, trimethoprim-sulfamethoxazole, and vancomycin. Participants were instructed to use the most current NCCLS guidelines for susceptibility testing. In January 2005 the NCCLS was renamed the CLSI, and its guidelines were implemented (2).

A performance target of 90% agreement for serotyping, oxacillin screening, and MIC determination was established. Evaluation criteria were consistent with phase I with the following modifications. (i) In 2001, comparison of MIC data to the modal MIC for each antibiotic-organism combination was added. The MIC results from all four laboratories (NCS and three QCC) were used to calculate the modal MIC. A variance of $\pm$1 $\log_2$ dilution was considered acceptable. (ii) In 2002, evaluation of the oxacillin zone size was added for all four laboratories (NCS and three QCC). An acceptable result was $\pm$3 mm of the mean zone size reported for each isolate, based on the observation that 1 $\log_2$ dilution in an MIC is approximately equal to 3 mm in a disk diffusion test (5).

The validation process was revised slightly from phase I to include submission of all nontypeable pneumococci, 15% of isolates that were "typeable" and 25% of the isolates that were resistant to any antibiotic routinely tested by the participating laboratories for MIC verification, using randomized selection criteria. An annual shipping schedule was provided to ensure the regular submission of isolates. All shipping followed International Air Transport Association regulations. Evaluation criteria were consistent with those used for phase I. A follow-up form, developed to track serotyping and MIC discrepancies, was initiated by the NCS when discrepant validation results were reported. The QCC were required to repeat the testing to resolve the discrepancy and submit the results back to the NCS within a specified time frame. Isolates were resubmitted if the problem was not resolved.

RESULTS

Phase I: EPT. Between October 1994 and August 1996, six QAPs (containing 30 isolates representing 11 vaccine-related and 7 non-vaccine-related serotypes) were distributed to the six participating countries. One country submitted results for only three of the six QAPs. A summary of EPT results for each laboratory is presented in Table 1.

Serotyping. All participants used the pool antisera typing system (14) supplemented with factor sera specific for the most

TABLE 1. Phase I and phase II EPT summary

Laboratory	% Agreement with the expected result									
	Serotyping		Oxacillin interpretation[a]		Oxacillin zone[b]		MIC agreement[c]		MIC interpretation[d]	
	I	II	I	II	I	II	I	II	I	II
A	83.3	96.9	93.3	97.7	NA	98.8	83.2	95.1	87.2	96.9
B	83.3	95.3	93.3	99.1	NA	86.7	78.5	85.6	85.8	93.4
C	93.3	90.0	96.7	96.2	NA	82.5	85.0	92.2	96.3	95.5
D	93.3	NA[e]	97.0	NA	NA	NA	84.0	NA	92.6	NA
E	83.3	NA	96.7	NA	NA	NA	85.7	NA	89.0	NA
F	91.7	NA	100	NA	NA	NA	81.3	NA	83.9	NA
Avg	88.0	93.8	96.1	97.6		89.3	83.0	91.0	89.1	95.3

[a] Interpretation of oxacillin zone to detect reduced susceptibility to penicillin based on a zone size of ≤19 mm.
[b] Reported zone size within 3 mm of the mean.
[c] MIC agreement within ±1 $\log_2$ dilution of the expected result.
[d] Susceptible, intermediate, or resistant interpretation category based on NCCLS/CLSI criteria that were current at the time of testing.
[e] NA, not assessed.

common serotypes. The overall serotyping correlation for the six QAPs was 88.0%. The serotype correlation for each QAP ranged between 75.8 to 100%. Although performance variation between participants was observed, individual laboratory proficiency was consistent over time. Discrepancies between the expected serotype and the reported result were randomly distributed and, in at least two instances, appeared to be related to incorrect numbering of the isolates by the participating laboratory. Three factoring errors were reported (incorrect serotype within the correct serogroup): one within group 23 and two within group 19.

Oxacillin. The six QAPs included 12 isolates that were expected to produce a zone of ≤19 mm; six produced no zone of inhibition (all with intermediate or resistant penicillin MICs), and six produced zone sizes of 10 to 17 mm. Three of these had intermediate penicillin MICs, and three were susceptible (MIC ≤ 0.06 μg/ml). The remaining 18 isolates were penicillin susceptible, with oxacillin zone sizes of 25 to 35 mm. Overall, 96.1% of the reported oxacillin zone sizes correctly detected strains with reduced penicillin susceptibility. The accuracy for the detection of reduced penicillin susceptibility for six strains with 10 to 17 mm zone sizes was 77.7% (six falsely susceptible results [very major errors]). Only 1 of the 30 isolates produced a discrepant result in more than one laboratory. The expected zone size for this isolate was 16 mm; the reported zones ranged from 18 to 27 mm.

MIC. The range of antibiotics tested by MIC in each of the participating laboratories varied. All were testing penicillin in October 1994. Testing additional antibiotics was gradually implemented over the next 12 months. One laboratory was not testing trimethoprim-sulfamethoxazole, and two were testing vancomycin by disk diffusion only.

Overall, average MIC correlation (±1 $\log_2$ dilution) for all participants for the six QAP was 83.0%. The performance of each participating laboratory ranged between 78.5 and 85.7%. The MIC correlation varied by antibiotic (Table 2). Poor correlation observed for erythromycin (68.7%) was associated with the MIC method; the laboratories using Etest consistently reported MICs that were higher than the BMD MIC reported by the NCS. Laboratories that were using BMD showed improved performance over time, while those that implemented Etest showed a trend toward poorer performance, attributable primarily to erythromycin. The laboratory using agar dilution showed consistent performance throughout phase I and reported MICs that were comparable to the BMD method.

The average interpretive category agreement for all QAP distributions was 89.1%; participant averages ranged from 83.9 to 96.3% with consistent performance over time. The average category agreement by antibiotic is presented in Table 2. A total of four major errors were reported, all occurring with chloramphenicol. There were 45 minor errors, 66.7% of which (30 of 45) resulted from a difference of 1 $\log_2$ dilution for an MIC close to the susceptible/intermediate/resistant breakpoint. There were no very major errors.

Phase I: validation process. Between August 1994 and November 1998 the NCS received 1,151 isolates for verification of serotype and/or MIC. The number of shipments and the number of isolates submitted varied by participant. Contaminated and nonviable samples reduced the total number of isolates available for evaluation (Table 3).

Serotype. There were 938 isolates available for serotype comparison over the 5-year period. Overall, correlation with the NCS result was 81.9% with wide participant variation.

TABLE 2. MIC and interpretive category agreement by antibiotic for EPT phase I and phase II

Antibiotic	% Agreement with the expected result			
	MIC[a]		Interpretive category[b]	
	I	II	I	II
Penicillin	82.7	93.3	88.7	93.6
Cefotaxime	81.3	96.0	93.4	92.7
Ceftriaxone	93.8	97.2	82.1	93.8
Chloramphenicol	86.6	95.0	93.3	96.2
Erythromycin	68.7	88.5	100	96.8
Trimethoprim-sulfamethoxazole	86.0	92.2	83.5	94.7
Vancomycin	85.6	98.7	NA[c]	NA

[a] MIC agreement within ±1 $\log_2$ dilution of the expected result.
[b] Susceptible, intermediate, or resistant interpretation category based on NCCLS/CLSI criteria that were current at the time of testing.
[c] NA, not assessed (susceptible category only).

TABLE 3. Phase I validation process results summary for six participating laboratories

Laboratory	No. of shipments	No. of isolates	No of isolates with MIC data (% of total)	Contamination rate (%)	Viability rate (%)	Serotype (%)[a]	MIC (%)[b]	Interpretive category (%)[c]
A	11	191	47 (24.6)	0.01	99.5	88.4%	69.4	76.5
B	3	26	6 (53.8)	0	100	96.2%	63.0	82.6
C	7	223	63 (28.3)	0.9	94.6	84.7%	87.4	94.1
D	11	358	21 (5.9)	11.2	69.8	79.4%	66.7	88.9
E	6	113	34 (30.1)	6.2	90.3	82.4%	85.7	80.2
F	6	240	15 (6.3)	2.5	96.3	74.0%	92.3	92.3
Total	44	1,151	194 (16.9)					
Avg				5.0	87.7	81.9%	80.4	85.8

[a] That is, the serotype reported by the participant correlates with the NCS serotype result.
[b] MIC agreement within ± 1 $\log_2$ dilution of the expected result as determined by the NCS.
[c] Susceptible, intermediate, or resistant interpretation category based on NCCLS/CLSI criteria that were current at the time of testing.

There was no obvious performance trend by laboratory over time except for a decline in performance for one laboratory that withdrew from the program in 1996 (data not shown). Discrepancies were randomly distributed among serotypes.

MIC and category agreement. Susceptibility data were submitted for 194 of 1,151 isolates received; 736 MICs were available for comparison. Participants achieved overall agreement of 80.4%, with wide interlaboratory performance variation that was also observed for category agreement. There were a total of 69 minor errors, 24 major errors, and 4 very major errors. The MIC validation process was impacted by the number of isolates for which MICs were submitted, compromising the evaluation of some participants.

Evaluation of phase I EQA model. In 1998, the EQA program was reviewed to assess and improve the process. Participation in EPT was excellent; however, gaps were noted, including the need for clear performance targets and a formal follow-up process for discrepant results to facilitate effective problem solving.

The validation process proved more challenging, with issues related to compliance with the original protocol. Unexpected variations in MIC testing methodology and inconsistent submission of isolates were of concern. Not all laboratories were forwarding the required proportion of isolates, and very few strains with reduced penicillin susceptibility were received from some regions with known penicillin resistance. High contamination rates and poor viability in some laboratories had the potential to compromise valuable organism collections.

International shipping of isolates for both EPT and the validation process was a major problem for some participants. At that time, all bacterial isolates were classified as "infectious agents" for transportation purposes, and many courier companies were reluctant to handle the shipments. Some countries required import permits and government regulations changed without notice, resulting in delayed delivery. The high cost associated with courier and customs fees was also of concern.

An important factor in the program evaluation was the growth of the SIREVA network from 6 to more than 20 countries representing almost all of Central and South America (Table 4). Tailoring one EQA program to meet the various technical needs of this large group was very difficult, as well as extremely costly if we were to continue to coordinate the program from Canada. A model that would move the primary EQA role into Latin America was required. This would mitigate the shipping problems, at least to and from Canada, better utilize limited funding by directing it into Latin America, and support the original goal of establishing regional self-sufficiency.

Three subregional QCC were selected from the original six participating countries based on operational capacity, geographic location, international shipping access for transport of isolates, national support for the program, local participation in the program, and proven ability and willingness to follow the EQA protocol. Each QCC would coordinate the EQA program for a specified group of countries, as well as provide technical support and regional training as required. The NCS would coordinate the EQA program for the three QCC. In 2005, the number of subregional QCC was reduced to two.

Phase II: EPT. Thirteen EPT events, including a total of 130 isolates representing 23 vaccine-related and 26 non-vaccine-related serotypes, were completed between June 1999 and December 2005. The QAPs were distributed to three QCC between 1999 and 2004 and to two QCC in 2005. A summary of EPT results for each laboratory is presented in Table 1.

Serotyping. Overall, serotyping correlation throughout the 6-year period averaged 93.8%. From 2003 onward all participants met or exceeded the 90% performance target, and in four EPT events all QCC achieved 100% correlation. Incorrect serotypes were reported for 16 of the 130 isolates and were randomly distributed. Six were factoring errors occurring within six different serogroups.

TABLE 4. SIREVA Network participating countries for phase I and phase II

Phase	Period (yr)	Participating countries
I	1994–1998	Argentina, Brazil, Chile, Colombia, Mexico, and Uruguay
II	1999–2005	Argentina, Brazil, Chile, Colombia, Mexico, Uruguay, Bolivia, CAREC[a], Costa Rica, Cuba, Dominican Republic, Ecuador, El Salvador, Guatemala, Honduras, Nicaragua, Panama, Paraguay, Peru, and Venezuela

[a] Caribbean Epidemiology Centre.

TABLE 5. Phase II validation process results summary for three participating laboratories

Laboratory	No. of shipments	No. of isolates	No of isolates with MIC data (% of total)	Contamination rate (%)	Viability rate (%)	Serotype[a] (%)	MIC[b] (%)	Interpretive category (%)[c]
A	10	338	290 (85.8)	0.6	98.5	95.5	95.0	96.0
B	4	116	89 (76.7)	1.7	90.5	77.5	77.2	87.5
C	11	689	316 (45.9)	0	97.7	85.4	91.2	94.6
Total	25	1,143	695 (60.8)					
Avg				0.3	97.2	88.1	90.5	94.3

[a] That is, the serotype reported by the participant correlates with NCS serotype result.
[b] MIC agreement within ± 1 $\log_2$ dilution of the expected result as determined by the NCS.
[c] Susceptible, intermediate, or resistant interpretation category based on NCCLS/CLSI criteria that were current at the time of testing.

Oxacillin. The 13 QAP included 81 penicillin-susceptible isolates with oxacillin zones of $\geq$20 mm and 49 isolates that failed the oxacillin screen (zone size $\leq$ 19 mm), including 27 with no zone of inhibition around the disk and 22 with a zone size of 7 to 19 mm. The overall average for correct interpretation (susceptible or reduced susceptibility to penicillin) was 97.6%. There were five falsely susceptible and four falsely resistant results reported. All falsely susceptible results occurred with strains with zone sizes of 7 to 19 mm. There was 92.1% correlation for the accurate detection of reduced penicillin susceptibility for the 22 isolates with a measurable, reduced zone size (i.e., 7 to 19 mm). In total, 8 of 130 isolates produced a discrepant interpretation in one or more QCC; 6 of these had "borderline" zone sizes of 15 to 22 mm.

Quantitative evaluation of the oxacillin zone size reported by the four participating labs was introduced in 2002. Overall, 91.4% of the results correlated within ± 3 mm of the average zone size (89.3% for the QCC only); however, there was wide variation in laboratory performance. Twenty-five discrepant results were reported, and two produced a change in interpretive category from reduced penicillin susceptibility to susceptible (very major errors). A pattern of consistently large zone sizes was observed for the QCC showing the poorest correlation.

MIC. All QCC were testing penicillin, cefotaxime or ceftriaxone, chloramphenicol, erythromycin, trimethoprim-sulfamethoxazole, and vancomycin using the BMD method. The overall agreement within 1 $\log_2$ dilution of the NCS MIC result for 1999 and 2000 and within 1 $\log_2$ dilution of the modal MIC for 2001–2005 was 91.0% for QCC only and 93.1% when NCS results were included in the analysis. All QCC showed improved performance over time and, from 2004 onward, all participants met or exceeded the 90% performance target. MIC correlation varied by antibiotic (Table 2). The poorest correlation was with erythromycin and was associated with a trend toward elevated MICs in one QCC and lower MICs in another during the early part of phase II.

The average interpretive category agreement for phase II QAP distributions was 95.3% with little variation by antibiotic (Table 2). There were 68 minor errors reported; 40 (58.8%) of these resulted from an MIC difference of only 1 $\log_2$ dilution between the reported and the expected result. Major errors occurred most frequently with chloramphenicol (15 of 24 [62.5%]); all resulted from a reported MIC of 8 μg/ml for strains with an expected MIC of 2 to 4 μg/ml. Seven major errors occurred with erythromycin. Three very major errors were reported by one QCC; all for the same strain.

Phase II: validation process. Between December 1999 and December 2005 the NCS received 25 shipments from three QCC, containing a total of 1,143 isolates. Participation in the validation process varied over the 6-year period. One QCC submitted isolates only during 2001 and 2002. Details of the validation results are presented in Table 5.

Serotyping. The overall serotype correlation of 88.1% was improved versus phase I (81.9%) but varied among the three QCC. Correlation varied for each shipment of isolates: lab A, 86.4 to 100%; lab B, 53.3 to 100%; and lab C, 73.9 to 94.3%. There was no obvious trend for individual QCC performance over the 6-year period. Discrepancies were randomly distributed among serotypes and were resolved on repeat testing by the submitting laboratory for 81% of the isolates. Discrepancies that were not resolved were assumed to be due to incorrect data submission (clerical error) or submission of the wrong isolate for validation.

MIC and category agreement. Susceptibility data were submitted for 695 of the isolates received; 4,105 MICs were available for comparison. Overall, participants achieved 90.5% correlation within 1 $\log_2$ dilution of the NCS MIC, although there was wide variation in QCC performance. Discrepancies were resolved on repeat testing by the submitting laboratory for 67.7% of the results (QCC range, 47.5 to 82.8%). Discrepancies that were not resolved were assumed to be due to incorrect data and/or isolate submission. Category agreement based on the submitted MIC for all participants averaged 94.3%. There were a total of 182 minor errors, 48 major errors, and 7 very major errors.

DISCUSSION

Published descriptions of international EQA programs for *S. pneumoniae* are very limited (1, 9, 13, 15) and describe the distribution of challenge isolates to participating laboratories with evaluation of results by the testing agency. To our knowledge, our report is the first description of a two-part model for validating laboratory results and also the first microbiology EQA program designed to support laboratory-based surveillance in developing countries.

We believe that establishing specific goals to guide the SIREVA project at the outset, combined with ensuring that all participants shared a clear understanding of the EQA princi-

ples, contributed to continued participation and international commitment to this project. Maintaining national ownership of isolates and data was an important principle for SIREVA partners.

One of the limitations of our EQA system was the difficulty associated with international transportation of isolates, especially in phase I. We are not the first group to encounter this problem (15). In addition, the successful implementation and resulting growth of the SIREVA project in Latin America produced unanticipated EQA management challenges that prompted the decentralization of the original EQA model.

Comparison of the results of our SIREVA EQA program is compromised by a lack of similar published international experience. We could find only one report of laboratory performance for pneumococcal serotyping (9). The overall correlation with the expected serotype for that study was 95%, with laboratory performance ranging from 90 to 100%. This performance is similar to our EPT phase II results and only slightly better than phase I. However, direct comparison of our EPT data with the Konradsen study may not be valid since our results reflect performance measurement over time rather than evaluation of a single event. Accurate serotyping is necessary to target and monitor vaccination programs and to measure serotype replacement in postvaccination populations. Correct serotype assignment within the serogroup (factoring) is particularly important since not all serotypes are included in the currently available vaccines (9). In both phases of our EPT, we observed relatively few "factoring" errors overall, indicating a high degree of proficiency with this technically difficult procedure.

The need to be able to accurately detect emerging antibiotic resistance, specifically for *S. pneumoniae*, and the need for enhanced proficiency testing have been identified by ourselves and by others (15). Accurate testing and interpretation of the oxacillin screen to detect reduced susceptibility to penicillin is critical for front line laboratories. The results from both phases of our EPT program indicate excellent performance. This compares favorably with two international laboratory proficiency testing exercises (1, 15). Each survey distributed a single pneumococcal isolate with expected zone sizes of 14 to 17 mm and 13 to 15 mm, respectively. Only 64 to 71% of the participating laboratories correctly classified these isolates as having reduced penicillin susceptibility. Performance for both phases of our EPT exceeded these results. One well-recognized North American proficiency testing program reported 99% correlation with the expected oxacillin screen result; however, this survey used a fully penicillin-resistant strain that was expected to give little or no zone of inhibition around the oxacillin disk (5). Interestingly, even in this setting, reported zone sizes ranged from 6 mm (no zone) to 20 mm. These reports and our experience suggest that isolates that produce a reduced but measurable (7 to 19 mm) oxacillin zone size may be more technically challenging for correct interpretation than those that produce no zone of inhibition around the disk.

The addition of quantitative analysis of the oxacillin zone sizes for phase II EPT enabled all QCC and the NCS to objectively evaluate performance. Using this tool we were able to demonstrate a trend toward larger zone sizes associated with occasional failure to detect reduced penicillin susceptibility by one QCC.

Comparing susceptibility results between laboratories is difficult unless standardized methodology is used. The SIREVA program was developed from a training model that provided standardized procedures and reagents. In spite of this, only three of the six phase I participants were routinely using the NCCLS/CLSI BMD method consistent with the initial protocol. Our observation of elevated erythromycin MICs associated with Etest has been reported by others and is likely due to the effect of carbon dioxide on the pH of the medium surface resulting in decreased erythromycin activity (5). All phase II participants used the BMD MIC method consistent with NCCLS/CLSI guidelines and obtained much-improved erythromycin correlation.

The occurrence of major errors for chloramphenicol is not surprising since a change in the interpretation between susceptible and resistant is based on an MIC difference of only 1 $\log_2$ dilution. In our experience, strains that produce an MIC of 8 μg/ml, resistant according to NCCLS/CLSI guidelines, are usually susceptible on repeat testing (MIC $\leq$ 4 μg/ml). We recommend that laboratories verify a chloramphenicol MIC of 8 μg/ml before reporting a resistant interpretation.

Most EQA challenges for *S. pneumoniae* have focused on accurate MIC testing for penicillin (1, 5, 13, 15). Excellent results for correct interpretive category were achieved when strains with very low (MIC $\leq$ 0.01 μg/ml) or very high (4 μg/ml) penicillin MICs were used (1, 5). However, the use of a borderline strain (MIC = 0.06 μg/ml) produced more variable results, with category agreement of only 80% (15). In a European study, seven isolates with MICs of 0.12 to 1.0 μg/ml were used to evaluate participant ability to correctly classify the strains as having reduced susceptibility to penicillin (13). An overall category agreement of 90.5% was achieved, with participant performance ranging from 78 to 100%. We achieved similar results in phase I, with improved performance in phase II demonstrating a high level of testing accuracy for this drug.

We could find only one published *S. pneumoniae* EQA survey that used a variation of ± 1 $\log_2$ dilution of the expected MIC as a measurement of accuracy (5). In that report, a single pneumococcal isolate was distributed to more than 4,200 laboratories in the United States. Susceptibility testing methodology and MICs for 14 antibiotics were reported. If MIC agreement for each antibiotic is calculated from their data, participants achieved the following MIC correlation: penicillin, 86%; cefotaxime, 87.4%; ceftriaxone, 87.6%; chloramphenicol, 92.9%; erythromycin, 48.6%; trimethoprim-sulfamethoxazole, 90%; and vancomycin, 94.9%. Our correlation for phase I is slightly lower for five of the seven antibiotics we tested. The phase II results, however, exceed those reported in the U.S. survey and show the high level of accuracy achieved for all antibiotics that we were evaluating in our EPT program.

Most laboratories do not handle EPT surveys in a blinded fashion, and a potential bias is introduced when samples are processed with increased attention that may not be typical of "routine" testing. The validation process we used offers a quality measurement that may be more representative of routine laboratory performance. It is not surprising that our data show a trend toward less accuracy for the validation process compared to the EPT, especially in phase I. Our improved validation process results for phase II demonstrate enhanced performance for routine laboratory testing.

We recognize that generally improved EQA performance observed in phase II compared to phase I may be due to a reduced number of participating laboratories, as well as the experience gained by the three laboratories selected for phase II. Regardless, the results achieved validate technical proficiency and support the subregional model of QCC that was implemented in 1999. The model has delivered quality tools that will continue to benefit the regional laboratory infrastructure. We believe that the SIREVA project has successfully reached its goal of creating a regional network of strong, self-sufficient laboratories that are able to lead local, national, and international EQA activities. The *S. pneumoniae* EQA program described here has continued into 2007, and recently the QCC have expanded the model to include *Haemophilus influenzae* and *Neisseria meningitidis*, demonstrating its usefulness for other international laboratory-based surveillance programs targeting vaccine-preventable bacterial disease.

ACKNOWLEDGMENTS

The members of the SIREVA Study Group include Gregory J. Tyrrell and Jean Weekes from the National Centre for Streptococcus, Edmonton, Alberta, Canada; Maria Luisa L. S. Guerra and Angela P. Brandão, Instituto Adolfo Lutz, São Paulo, Brazil; Olga Marina Sanabria and Maria Victoria Ovalle, Instituto Nacional de Salud, Bogotá, Colombia; Julio Pace, Sofia Fossati, Paula Gagetti, and Elsa Chavez, Instituto Nacional de Enfermedades Infecciosas ANLIS Carlos Malbrán, Buenos Aires, Argentina; Juan Carlos Hormazabal and Cecilia Carmona, Instituto de Salud Pública, Santiago de Chile, Chile; María Elena Velázquez-Meza, Noemí Carnalla-Barajas, and Margarita Hernández-Salgado, Instituto Nacional de Salud Pública, Cuernavaca, Mexico; Fortino Solórzano-Santos, Hospital de Pediatría, Centro Médico Nacional, Siglo XXI, Instituto Mexicano del Seguro Social, Mexico City, Mexico; and Gabriela Garcia Gabarrot and Gabriel Perez, Departamento de Laboratorios de Salud Pública, Montevideo, Uruguay.

REFERENCES

1. **Chaitram, J. M., L. A. Jevitt, S. Lary, F. C. Tenover, et al.** 2003. The World Health Organization's external quality assurance system proficiency testing program has improved the accuracy of antimicrobial susceptibility testing and reporting among participating laboratories using NCCLS methods. J. Clin. Microbiol. **41:**2372–2377.
2. **Clinical and Laboratory Standards Institute.** 2005. Performance standards for antimicrobial susceptibility testing; 15th informational supplement. CLSI document M100-S15. CLSI, Wayne, PA.
3. **Di Fabio, J. L., A. Homma, and C. de Quadros.** 1997. Pan American Health Organization epidemiological surveillance network for *Streptococcus pneumoniae*. Microb. Drug Resist. **3:**131–133.
4. **Di Fabio, J. L., E. Castañeda, C. I. Agudelo, F. de la Hoz, M. Hortal, T. Camou, G. Echániz-Aviles, M. Noemi, C. Barajas, I. Heitmann, J. C. Hormazabal, M. C. Brandileone, V. Simonsen dias Vieira, M. Regueira, R. Ruvinski, A. Corso, M. Lovgren, J. A. Talbot, C. de Quadros, et al.** 2001. Evolution of *Streptococcus pneumoniae* serotypes and penicillin susceptibility in Latin America, Sireva-Vigía Group 1993–1999. Pediatr. Infect. Dis. J. **20:**959–967.
5. **Doern, G. V., A. B. Brueggemann, M. A. Pfaller, and R. N. Jones.** 1999. Assessment of laboratory performance with *Streptococcus pneumoniae* antimicrobial susceptibility testing in the United States. Arch. Pathol. Lab. Med. **123:**285–289.
6. **Elder, B. L., S. A. Hansen, J. A. Kellog, F. J. Marsik, and R. J. Zabransky.** 1997. Cumitech 31: verification and validation of procedures in the clinical microbiology laboratory. American Society for Microbiology, Washington, DC.
7. **Facklam, R. R., and J. A. Washington.** 1991. *Streptococcus* and related catalase-negative gram-positive cocci, p. 238–257. *In* A. Balows, W. J. Hausler, Jr., K. L. Herrmann, H. D. Isenberg, and H. J. Shadomy (ed.), Manual of clinical microbiology, 5th ed. American Society for Microbiology, Washington, DC.
8. **Kertesz, D. A., J. L. Di Fabio, M. C. C. Brandileone, E. Castañeda, G. Echániz-Aviles, I. Heitmann, A. Homma, M. Hortal, M. Lovgren, R. O. Ruvinski, J. A. Talbot, J. Weekes, J. S. Spika, et al.** 1998. Invasive *Streptococcus pneumoniae* infection in Latin American children: results of the Pan American Health Organization surveillance study. Clin. Infect. Dis. **26:**1355–1361.
9. **Konradsen, H. B.** 2005. Validation of serotyping of *Streptococcus pneumoniae* in Europe. Vaccine **23:**1368–1373.
10. **Lund, R., and J. Henrichsen.** 1978. Laboratory diagnosis, serology and epidemiology of *Streptococcus pneumoniae*, p. 241–262 *In* T. Bergan and J. R. Norris (ed.), Methods in microbiology, vol. 12. Academic Press, Inc., New York, NY.
11. **National Committee for Clinical Laboratory Standards.** 1993. Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically, 3rd ed.; approved standard. NCCLS document M7-A3. National Committee for Clinical Laboratory Standards, Wayne, PA.
12. **National Committee for Clinical Laboratory Standards.** 1993. Performance standards for antimicrobial disk susceptibility tests, 5th ed.; approved standard. NCCLS document M2-A5. National Committee for Clinical Laboratory Standards, Wayne, PA.
13. **Snell, J. J. S., and D. F. J. Brown.** 2001. External quality assessment of antimicrobial susceptibility testing in Europe. J. Antimicrob. Chemother. **47:**801–810.
14. **Sørensen, U. B. S.** 1993. Typing of pneumococci by using 12 pooled antisera. J. Clin. Microbiol. **31:**2097–2100.
15. **Tenover, F. C., M. J. Mohammed, J. Stelling, T. O'Brien, and R. Williams.** 2001. Ability of laboratories to detect emerging antimicrobial resistance: proficiency testing and quality control results from the World Health Organization's external quality assurance system for antimicrobial susceptibility testing. J. Clin. Microbiol. **39:**241–250.