THREE YEARS OF COLLABORATION ON THE STANDARDIZATION OF CHAGAS' DISEASE SERODIAGNOSIS IN THE AMERICAS: AN APPRAISAL[1]

Mario E. Camargo,[2] Elsa L. Segura,[3] Irving G. Kagan,[4] José Maria Pacheco Souza,[5] José da Rocha Carvalheiro,[6] Jorge F. Yanovsky,[7] and Maria Carolina S. Guimarães[8]

INTRODUCTION

Disagreement between Chagas' disease serodiagnoses performed by different laboratories is a common problem. Indeed, because of variations in technical procedures and reagents, as well as in criteria for evaluating the results of serologic tests, discrepancies between the results obtained by different laboratories are not really unexpected. For this reason, a collaborative study was begun in 1981 that was directed at developing a continental standardization program for Chagas' disease serodiagnosis; the study was sponsored by the United Nations Development Program/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases.

This report outlines the progress of the study and work performed through October 1983, presents some statistical data, and describes the results obtained. It also comments on the experience gained and presents a number of proposals for future action.

[1] This article will also be published in Spanish in the *Boletín de la Oficina Sanitaria Panamericana*, 1987. The investigation reported here received financial support from the United Nations Development Program/World Bank/World Health Organization Special Program for Research and Training in Tropical Diseases.

[2] Laboratory of Immunology and Seroepidemiology, Instituto de Medicina Tropical de São Paulo, Av. Dr. Eneas de Carvalho Aguiar 470, São Paulo, 05403, São Paulo, Brazil.

[3] Director, Instituto Nacional de Diagnóstico e Investigación de la Enfermedad de Chagas "Dr. Mario Fatala Chaben," Paseo Colón 568, Buenos Aires, Argentina.

[4] Director, Division of Parasitic Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia, USA.

[5] Professor, Department of Epidemiology, Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, Brazil.

[6] Professor, Department of Social and Preventive Medicine, Faculdade de Medicina de Riberão Preto, Universidade de São Paulo, Riberão Preto, São Paulo, Brazil.

[7] Director, Polychaco Foundation, Buenos Aires, Argentina.

[8] Laboratory of Immunology and Seroepidemiology, Instituto de Medicina Tropical de São Paulo, Brazil.

MATERIALS AND METHODS

Preparation of Serum Samples

In order to compare diagnostic techniques, serum panels were exchanged by participating laboratories. A low-cost, practical procedure was developed to eliminate the need for expensive air-mailing of frozen sera or for lyophilized samples. This procedure consisted of preserving serum by mixing it with an equal volume of analytical grade glycerin, a technique found very satisfactory for purposes of maintaining serum titers unchanged during years of storage at − 20°C or during weeks or months of storage at room temperature.

Since small amounts of sera suffice for performing a variety of tests, 200 to 300 microliter aliquots of glycerin-preserved samples were distributed in 4-5 cm segments of plastic tubing about 3 mm in diameter made by Eletrovet (Eletro Veterinária Limitada, São Paulo, Brazil). These "straws" were sealed by introducing small polystyrene spheres (Araguaia Indústria e Comércio Limitada, São Paulo, Brazil) at both ends. Collections of 100 to 200 such "straws" could then be packaged in an ordinary letter and air-mailed.

We found that two beads should be used at each end of the straw to ensure against leakage. Also, a recently distilled glycerin was found preferable for preserving sera, since it appears that glycerin which has been stored a long time may contain decomposition products that can lead to an eventual decrease in serum reactivity. During shipping the straws were protected against excessive pressure by being placed in sufficiently deep grooves of a corrugated cardboard sheet that was covered by a stiff cardboard plate.

The Three Comparative Laboratory Studies

A group known as the Continental Group for Studies on Chagas' Disease Serology, with representatives from nine countries, was established during a workshop on Chagas' disease serology sponsored by the WHO/UNDP/TDR program in São Paulo, Brazil, in July 1980. This group provided the basis for initiating comparative studies by three laboratories—one in Argentina,[9] one in Brazil,[10] and one in the United States[11]—as a preliminary step toward an expanded study involving laboratories in other countries.

These three laboratories performed two comparative studies. In both studies each laboratory examined two series (A and B) of serum samples, series A being provided by Argentina and series B by Brazil. Each series contained 150 sera; series A was collected from subjects with no clinical histories available, while series B was collected from two groups of subjects, one known to be infected with *T. cruzi* and the other known to consist mainly of high-income residents in a nonchagasic area who were presumably negative for *T. cruzi* infection but who were probably not representative of the populations living in areas endemic for Chagas' disease. The first of the two

[9] The National Institute of Chagas' Disease Diagnosis and Research "Dr. Mario Fatala Chaben" (Instituto Nacional de Diagnóstico e Investigación de la Enfermedad de Chagas "Dr. Mario Fatala Chaben"), Buenos Aires, Argentina.

[10] The São Paulo Institute of Tropical Medicine (Instituto de Medicina Tropical de São Paulo), São Paulo, Brazil.

[11] The Division of Parasitic Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia, USA.

comparative studies employed the serologic test procedures regularly used in each of the three laboratories, while the second employed uniform protocols for each test and a uniform commercial reagent.

Following these two studies, the three participating laboratories were joined by four other laboratories in Bolivia, Chile, Colombia, and Panama for the purpose of conducting a third study.[12] This latter was performed with 200 serum samples from two groups—about 100 samples from each group. One group was composed of subjects deemed very likely on the basis of clinical, parasitologic, and epidemiologic data to be infected with *T. cruzi*, and the other was composed of subjects born and living in nonendemic areas of Brazil's Espírito Santo State under socioeconomic conditions comparable to those prevailing in the endemic areas. About half of the probably infected subjects—including both subjects with myocardiopathy and with the indeterminate form of the disease—were positive for *T. cruzi* by xenodiagnosis; the remainder of this group consisted of subjects who were seroreactive in various tests to *T. cruzi* and whose clinical and epidemiologic histories were compatible with Chagas' disease.

[12] This was done in coordination with Dr. C. la Fuente, Centro Nacional de Enfermedades Tropicales, Santa Cruz, Bolivia; Dr. H. Schenone, Departamento de Microbiología y Parasitología, Santiago, Chile; Dr. F. Guhl, Universidad de los Andes, Bogotá, Colombia; and Dr. O. E. Sousa, Facultad de Medicina, Universidad de Panamá, Panama.

DESCRIPTION AND RESULTS OF THE THREE STUDIES

The First Comparative Study

In this study, done in 1981, each laboratory performed its own tests, and the diagnosis was based upon the laboratory's own criteria for positivity. The Argentine laboratory tested the sera by hemagglutination (HA), immunofluorescence (IF), and direct agglutination (DA); the Brazilian laboratory employed HA, IF, and complement fixation (CF); and the US laboratory used HA, CF, and DA. At each laboratory a sample was considered positive when at least one of the tests employed was positive. Although a positive response to only one test was exceptional, this criterion was used because of the laboratories' different subjective ways of valuing the results of individual tests. (For example, one of the laboratories insisted on making a positive diagnosis whenever an isolated complement fixation test yielded positive results.)

For comparative evaluation of the laboratory diagnoses, Kappa statistics were used to measure agreement corrected for possible chance effects. (Kappa is 0 for mere chance agreement, reaches + 1 for total agreement, and shows negative values in cases where disagreement is more than merely casual—*1,2.*)

The results obtained with the serum panels (see Table 1) reflected not only differences in the reactivity of the tests employed by different laboratories, but also differences in the criteria used to score sera as positive or negative. It should be noted that this initial study intentionally placed no restrictions on the technical procedures or tests performed

TABLE 1. The degree of diagnostic agreement, on a scale of 0 to 1, obtained by the three laboratories participating in the first two studies. In each case the first column shows the degree of agreement without statistical adjustment for chance effects and the second column shows the Kappa-adjusted degree of agreement. AR indicates the Argentine laboratory, BR the Brazilian laboratory, and US the United States laboratory.

Comparison of results from indicated laboratories	First study agreement indices		Second study agreement indices	
	Crude	Adjusted	Crude	Adjusted
	Series A (Argentine) samples			
AR and BR	0.85	0.67	0.91	0.81
AR and US	0.79	0.56	0.85	0.70
BR and US	0.81	0.61	0.87	0.74
	Series B (Brazilian) samples			
AR and BR	0.85	0.69	0.95	0.91
AR and US	0.83	0.65	0.94	0.88
BR and US	0.93	0.85	0.96	0.92

at each laboratory. Table 1 shows the extent of crude agreement in this study and the degree of agreement after adjustment for chance effects. As may be seen, somewhat better agreement was obtained with the Brazilian (series B) sera than with the Argentine (series A) sera.

The Second Comparative Study

For the purpose of obtaining more uniform results, protocols were established for every test and a second study was performed in 1982 with two new panels of sera (A and B), prepared and distributed as in the initial study. As requested by the Steering Committee of the WHO/UNDP/TDR Chagas' Disease Scientific Working Group, tests with a commercial reagent were conducted in addition to the other tests, the reagent selected being one used for the hemagglutination test because this test was routinely performed in all three participating laboratories.

The results obtained with series A and series B, expressed as agreement indices, showed somewhat better agreement than did the results obtained in the first study (see columns 3 and 4 of Table 1). It was also found that by varying the cutoff titers used to define whether a given serum was responding positively or negatively, even closer agreement could be attained. (For example, by changing the cutoff titer in the DA test at one laboratory from 1:16 to 1:32, the specificity index improved, rising from 0.61 to 0.87, without causing any significant variation in test sensitivity.)

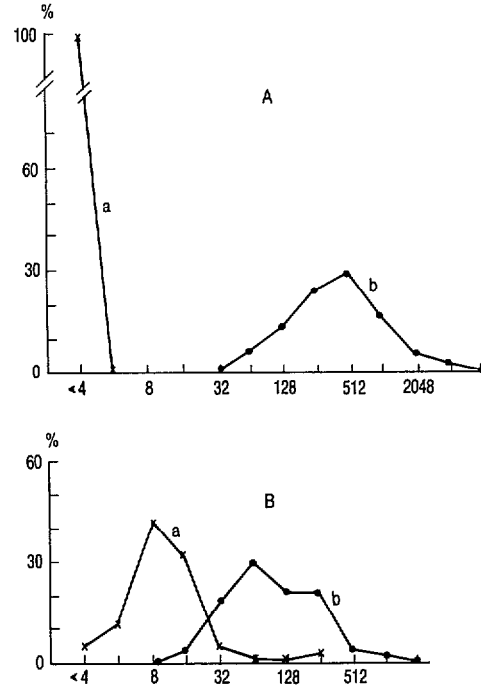
The Third Comparative Study

In 1983, panels containing the previously described collection of 200 coded serum samples, about 50% from Chagas' disease patients, were distributed to each of the participating laboratories. Those in Argentina, Brazil, and the United States performed the same tests that had been done in the second

study. The tests performed at the new laboratories were IF and HA in Bolivia, HA in Chile, IF and enzyme-linked immunosorbent assay (ELISA) in Colombia, and ELISA in Panama—all according to each laboratory's own protocol. The results of each test were then decoded, and distribution curves were constructed that showed the frequency with which different titers were obtained with sera from the chagasic and nonchagasic subjects.

The curves developed in this manner brought out several points. Among other things, they showed that different tests varied considerably in their ability to discriminate between sera from infected and uninfected subjects, even when the tests were performed in the same laboratory (Figure 1). They also pointed up differences between results obtained with the same test at different laboratories (Figure 2) and at different laboratories using different reagents (Figure 3), and provided a good overview of divergences produced by use of different tests at different laboratories—as exemplified by the divergences shown in Figure 4. The usefulness of distribution curves in evaluating test validity is indicated by the similarity of patterns obtained by the same tests at the same laboratories, even when they were conducted months apart with different serum panels. Such concordance is exhibited by the Figure 5 distribution curves derived from the second and third studies.

FIGURE 1. Distribution curves showing the frequency with which particular titers were obtained by (A) hemagglutination and (B) direct agglutination testing of 200 sera from (a) nonchagasic and (b) chagasic subjects at the same laboratory, using the same serum panel for both tests.

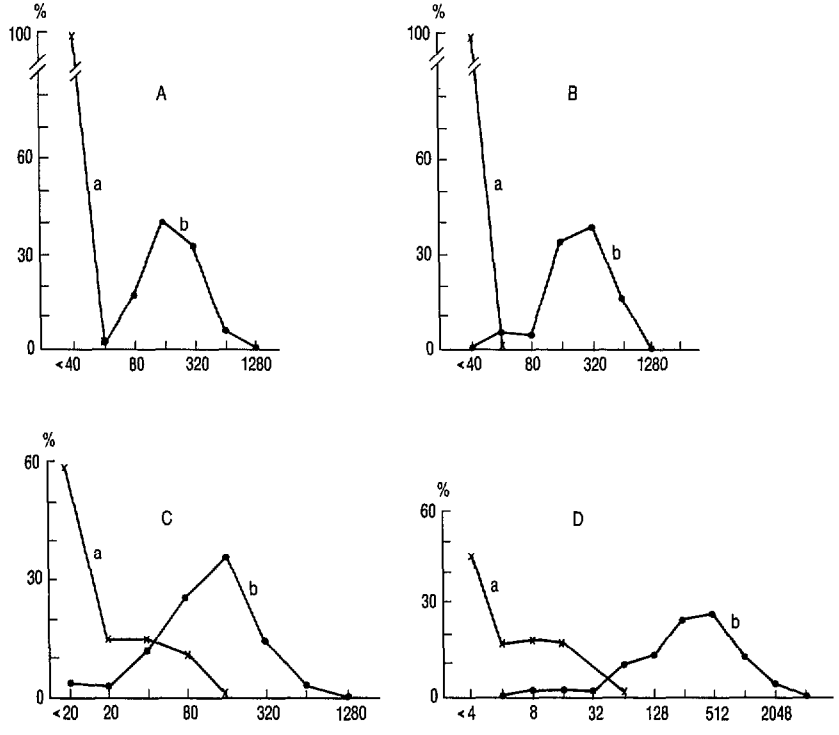


DISCUSSION

Many circumstances make it hard to standardize Chagas' disease serodiagnosis. A variety of tests based on different techniques are available, and the laboratories involved tend to select particular tests on the basis of personal preferences or the availability of equipment and reagents. The fact that different batches of reagents—or even successive batches from the same source—can vary in quality also adds to the potential for divergent, heterogeneous results.

In addition, parasite antigens may vary widely, from whole cells to

FIGURE 2. Distribution curves showing the frequency with which particular titers were obtained by immunofluorescence testing of 200 sera from (a) nonchagasic and (b) chagasic subjects at four different participating laboratories (A, B, C, and D), using the same serum panel at each laboratory.

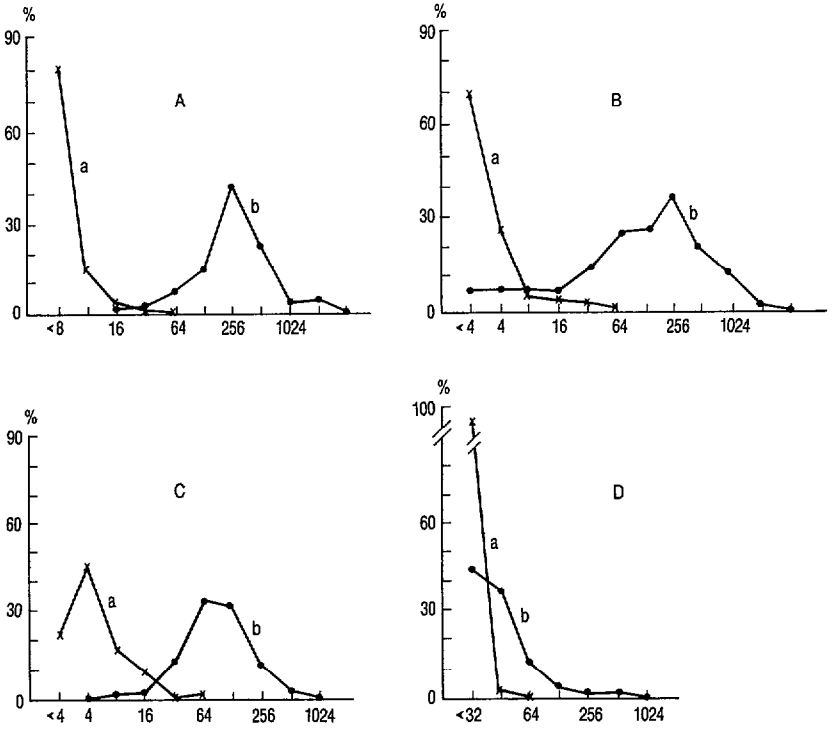


crude cell extracts to more or less purified fractions. It is noteworthy, however, that very close agreement (96.9% agreement in tests of over 10,000 sera) has been reported (*3*) in hemagglutination, complement fixation, immunofluorescence, and flocculation tests employing a variety of different antigen preparations. This suggests that the chief advantage of using defined *Trypanosoma cruzi* components, such as those described elsewhere (*4–7*), instead of less defined antigens (e.g., crude extracts or whole parasites) is that this helps to avoid cross-reactions like those frequently seen with (mainly visceral) leishmania infections.

The observation of differences between tests employing similar antigens, together with the previously reported close agreement between tests using very different antigen preparations, suggests strongly that test divergences are due primarily to factors interfering with the signalling mechanism that expresses the antigen-antibody reaction as agglutination, fixation of complement, fluorescence, etc.

In fact, it is known that signal-originating phenomena are prone to a variety of disruptive influences that tend to introduce "noise" or reduce signal intensity. For example, a poor antiglobulin conjugate may cause nonspecific fluorescence or weak staining of reactive sera; and in hemagglutination tests the nature of diluting solutions,

FIGURE 3. Distribution curves showing the frequency with which particular titers were obtained by hemagglutination testing of 200 sera from (a) nonchagasic and (b) chagasic subjects at four different participating laboratories (A, B, C, and D), using the same serum panel but different reagents at each laboratory.



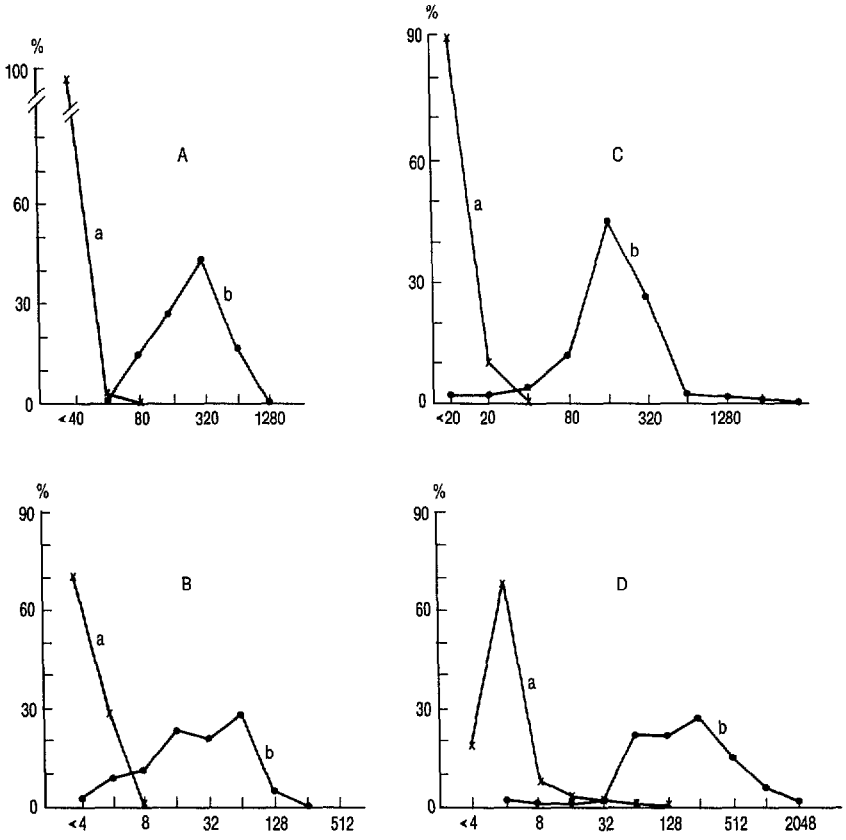
conditions affecting the cells' ability to agglutinate, and so forth can influence the tests' sensitivity and specificity. Nor are these isolated examples, for indeed many factors such as these can influence each kind of test employed.

It should also be noted that nonspecific positive responses can result from "natural" IgM antibodies reacting with *T. cruzi* antigens. These antibodies, which are found in varying degrees and frequencies among different populations (*8*), interfere mainly with tests such as immunofluorescence and direct agglutination that employ whole parasites as antigen.

In addition, regional differences in the antigenicity of local *T. cruzi* strains could create regional differences in test results. However, the use of complex antigens (represented by whole extracts or parasite bodies that are rich in components common to different *T. cruzi* strains), as in the work reported here, should minimize such differences.

Divergences of results between laboratories can be minimized by reporting the reactivity of sera relative to that of a standard positive serum. In the present study, it was observed that this normalization of titers in "units" relative to a standard positive reference serum

FIGURE 4. Distribution curves showing the frequency with which particular titers were obtained with 200 sera from (a) nonchagasic and (b) chagasic subjects by means of (A) direct agglutination test in Argentina, (B) hemagglutination test in the United States, (C) the enzyme-linked immunosorbent assay (ELISA) method in Colombia, and (D) the ELISA method in Panama. The serum panels tested contained sera from Brazilian subjects and were mailed to the testing laboratories from Brazil.

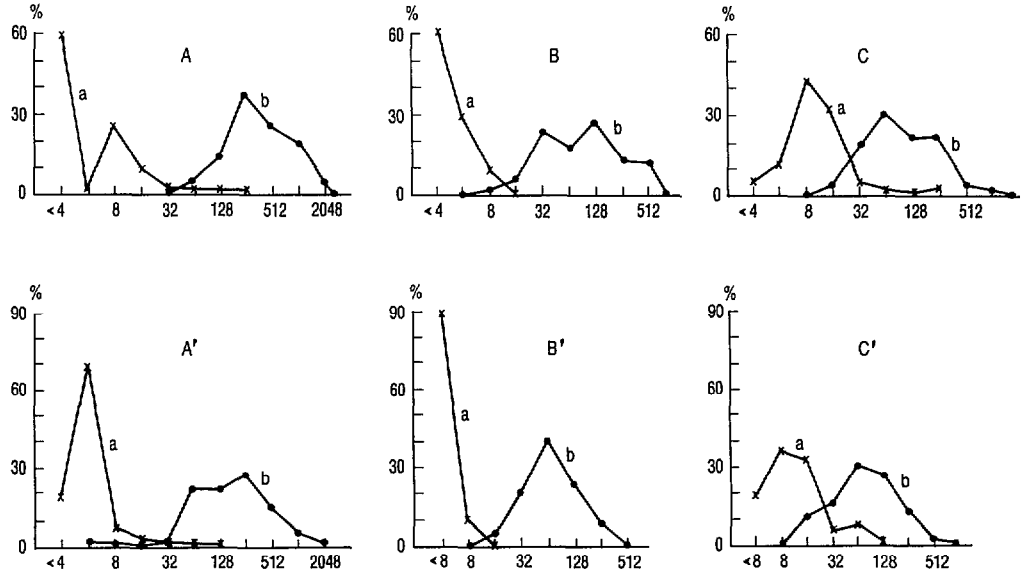


did not succeed in correcting diagnostic divergences. Nevertheless, such a procedure would seem to offer a helpful way of correcting for those variations that depend on differences in test sensitivity related to the intensity of the signal produced, as in the case of fluorescence.

It is evident that a rigorously standardized, sensitive, and specific test with defined antigen, reagents, and procedures could constitute a reference against which to evaluate other tests and reagents. Before such a standard is defined, however, serum panels from subjects well-diagnosed as being positive or negative for *T. cruzi* infection would appear to provide the best available standard.

The proposal to adopt a panel of reference sera containing anywhere from 100 to over 200 samples suggested a promising place to start. Accordingly, serologic collections of this type were made, using sera from subjects with chagasic and nonchagasic diagnoses defined

FIGURE 5. Distribution curves showing the frequencies with which particular titers were obtained using two different serum panels subjected to direct agglutination testing at laboratory A (charts A and A′), hemagglutination testing at laboratory B (charts B and B′), and direct agglutination testing at laboratory C (charts C and C′). The results shown in charts A′, B′, and C′ were obtained from the third study, several months after those shown in charts A, B, and C from the second study.



with the help of independent criteria including xenodiagnosis, clinical findings, and epidemiologic data. Although all the sera used in our studies came from only two countries (Argentina and Brazil), such panels could be improved in the future by including samples from other sources.

By and large, it would appear practical to characterize particular tests by assessing the reactivity obtained, expressed as serum titers, in testing a single panel of chagasic and nonchagasic sera. In comparing different tests, quantitative results are more illustrative than mere qualitative results based upon a previously established cutoff titer or level of reactivity. A convenient graphic image of each test can be obtained through distribution curves showing the frequency with which particular titers were obtained. In this way the discriminating quality of the test is clearly expressed, and causes of observed limitations can be studied. For example, charts C and D in Figure 2 (which compares immunofluorescence tests) indicate a likely need to improve the quality of the fluorescence conjugate. Similarly, chart D in Figure 3 (which compares hemagglutination tests) suggests that possible insufficient sensitization of erythrocytes should be investigated. It should also be noted that similar results can be attained on different occasions, even with panels containing samples from different subjects, and even when the tests are performed some months apart (see Figure 5).

In general, the stability of the glycerin-preserved serum panels used in

our studies appeared to be excellent. Figure 4 shows that highly satisfactory results were obtained with sera mailed from Brazil to test laboratories as far away as Argentina and Panama. Since low temperatures are not necessary for transportation of glycerin-serum mixtures, this stability makes it possible to exchange serum panels through the mail—a method that is practical, cheap, and free of troublesome customs delays.

Concluding Remarks

Up to now, laboratories conducting serologic tests for Chagas' disease have generally developed their reagents and tests according to their own experience and the particular characteristics of the population served—including the prevalence of infection within that population.

It is also clear that other means of Chagas' disease diagnosis—including xenodiagnosis and other parasitologic methods—have significant limitations. Nevertheless, evaluation of serologic methods for Chagas' disease diagnosis has generally been based on xenodiagnosis, combined with data from various clinical diagnostic procedures. As a result, the apparent sensitivity and specificity determined by the evaluation are not necessarily "true" values but are merely relative values measured in terms of the sensitivity and specificity of the other methods. Even a detailed comparison of the diagnostic results obtained for each subject yield only "co-positive" and "co-negative" values corresponding to the respective probability of obtaining positive or negative results by both procedures.

Thus, besides there being a diversity of tests (which is desirable) and a diversity of techniques (which is unavoidable), the main problem in Chagas' disease diagnosis relates to the reference criteria that were adopted to determine the sensitivity and specificity of the tests commonly used. In general, because of the interchanges between laboratories in the past, it seems likely that these tests are not independent but maintain among themselves complex and variable relations. However, no one can be sure that the tests will agree when applied to the same battery of sera by different laboratories. Indeed, comparison of the titers obtained in our studies indicates that absolute agreement is practically unattainable; and even analysis of the results in terms of the final diagnosis (whether sera were positive or negative) shows that co-positivity and co-negativity did not reach 100%, even when the optimum "cutoff" titer was selected.

Therefore, it appears that a good start on continental standardization of Chagas' disease serodiagnosis could be made by distributing identical serum panels, with the minimum acceptable sensitivity and specificity values precisely defined, to the main diagnostic laboratories in Latin America. These values, in fact, would merely be co-positivity and co-negativity indices, since only a probable diagnosis can be established for most patients. However, establishment of such parameters for this collection of sera should make it possible for the laboratories to reproduce a standard that is uniform.

To help ensure comparability, it is recommended that the tests be performed on the same standard serum panel. However, this standard panel

should be subject to modifications as the work extends through the Americas and new serum samples are introduced into the study. Of course, only future work can determine what modifications are appropriate; and so the composition of the panel should be dynamic, changing in such a way as to provide satisfactory comparability on a continental scale between laboratories working with a wide variety of differing local conditions.

Another important problem needing attention is the question of how well serologic testing for Chagas' disease yields public benefits. In general, the benefits derived depend on the predictive values, positive or negative, and hence the diagnostic accuracy, of the tests employed. Beyond that, once the validity of the tests can be assessed in terms of a standard serum panel tested by other laboratories, then the benefits to be derived will depend on the real prevalence of the disease among the population to be tested and the proposed objectives of the testing program.

Of course, intentional variations in test sensitivity and specificity can be established for different purposes—such as screening blood donors as compared to screening asymptomatic but possibly infected people who would need preventive treatment to avoid secondary clinical consequences—and in this way a better relationship between costs and benefits can be obtained. Also, test cut-off levels can be made to reflect local reality if they are established in accord with the local prevalence of infection and the distribution frequencies of titers for infected and uninfected individuals. However, none of these considerations detracts from the need for a common reference standard, and in fact there is every reason to believe that establishment of such a standard would yield important benefits for public health in Chagas-endemic zones.

With regard to the future, it should be noted that a fourth set of trials is now under development in the Americas, with the participation of a larger number of laboratories and employment of a defined panel of serum samples.

Summary

A collaborative study aimed at developing a continental standardization program for the serodiagnosis of Chagas' disease was begun in Argentina, Brazil, and the United States in 1981. Using two panels of sera, one collected in Argentina and one in Brazil, single laboratories in the three countries conducted serologic tests according to procedures and with reagents regularly used at those laboratories. A technique was developed to facilitate the interchange of serum samples. A second set of trials was then done, following uniform protocols. These trials, which tested a panel of sera from apparently infected and apparently uninfected subjects, included a hemagglutination test in which all three laboratories used the same commercial reagent.

These two sets of trials were followed by a third using a panel of sera with samples from subjects infected with *T. cruzi* and uninfected subjects—as diagnosed to a high degree of probability on clinical, parasitologic, and epidemiologic grounds. Four additional laboratories in Bolivia, Chile, Colombia, and Panama participated in these third trials.

The results of the first trial reflected noteworthy differences in the reactivity of the various serologic tests and also in the criteria used to score sera as positive or negative. The second trial showed somewhat better agreement between the three laboratories' results and indicated that by varying the cutoff titers used to define whether a given serum was positive or negative, even closer agreement could be attained.

The third trial brought out a variety of points. It demonstrated that the diagnostic value of different tests varied considerably, even when the tests were performed in the same laboratory. It also highlighted important differences between the results obtained with the same test at different laboratories and with different tests at different laboratories.

At present it is not possible to establish a perfect standard for diagnosis of chagasic infection because no completely sure diagnostic method exists. Therefore, a given test's apparent sensitivity and specificity are not necessarily its true ones but are rather its sensitivity and specificity measured relative to those of other methods. Thus, besides the diversity of available tests and techniques, a prime problem in diagnosing chagasic infections is the absence of firm reference criteria. Within this context, it appears that a good start on standardization of Chagas' disease serodiagnosis in the Americas could be made by distributing identical serum panels, with the minimum acceptable sensitivity and specificity values precisely defined, to the main diagnostic laboratories involved. This should make it possible for the participating laboratories to follow uniform standards; and such standards, by improving diagnostic accuracy and epidemiologic knowledge, would seem calculated to yield important benefits for public health in Chagas-endemic zones.

References

1 Fleiss, J. L., J. Cohen, and B. S. Everitt. Large simple standard error of Kappa and weight Kappa. *Psychol Bull* 72:323, 1969.

2 Cohen, J. A coefficient of agreement for nominal scale. *Educational and Psychological Measurement* 20:37, 1970.

3 Takei, K. Estudo da eficiência relativa dos diferentes testes sorológicos utilizados no diagnóstico da doença de Chagas: Resultados observados na análise de 10,181 soros. Doctoral thesis. Departmento de Microbiologia e Imunologia do Instituto de Ciências Biomédicas de Universidade de São Paulo, São Paulo, 1982.

4 Snary, D., and L. Hudson. *Trypanosoma cruzi* cell surface proteins: Identification of one major glycoprotein. *FEBS Lett* 100:166, 1979.

5 Araujo, F. G., and L. S. Remington. Characterization of stages and strains of *Trypanosoma cruzi* by analysis of cell membrane components. *J Immunol* 127:855–859, 1981.

6 Nogueira, N., S. Chaplan, J. D. Tydings, J. Unkeless, and Z. Cohn. *Trypanosoma cruzi*: Surface antigens of blood and culture forms. *J Exp Med* 153:629–639, 1981.

7 Scharfstein, J., M. M. Rodrigues, C. A. Alves, W. De Souza, J. O. Previato, and L. Mendonça-Previato. *Trypanosoma cruzi*: Description of a highly purified surface antigen defined by human antibodies. *J Immunol* 131:972–976, 1983.

8 Storni, P., F. J. Bolsi, and J. F. Yanovsky. Reacción de agglutinación direta para diagnóstico de la enfermedad de Chagas. Utilización sistemática del 2-mercapto-ethanol para la eliminación de las aglutininas inespecíficas. *Medicina (Buenos Aires)* 35:67, 1975.