V.D.R.L., EAGLE, PRICE, AND KOLMER TESTS
A COMPARISON USING 1,029 SOUTH AFRICAN BANTU SERA*

BY
S. B. GRIFFITHS†, G. C. BUCKLE, AND J. W. HILL
Natal Provincial Pathological Service and University of Natal

As no published results were available of valid evaluations of common serological reactions found amongst the various racial groups of South Africa, the serological department of the Natal Pathological Service initiated a comprehensive investigation of these reactions in the racial groups of Natal. It is hoped to publish the results in a series of papers. This paper reports the results of testing 1,029 apparently healthy Bantu males using the Eagle flocculation test (Eagle), the V.D.R.L. slide test (VDRL), the Kolmer complement-fixation test (Kolmer), and the Price precipitation reaction (Price).

Material and Methods

Subjects.—The subjects were young, apparently healthy Bantu males applying for work at the Department of Native Affairs of the Municipality of Durban. Blood was obtained by venepuncture from approximately the first 120 subjects in the queue on each of ten mornings. The only selection depended on obtaining those with large enough veins to allow the collection of sufficient blood quickly enough. All blood samples of less than 8 ml. were rejected.

Sera.—The sera were separated from the blood clots, inactivated for 30 min. at 56°C., and tested on the day of collection. Any surplus sera were stored frozen at −20°C.

Technique.—The techniques used for the VDRL, Eagle, and Kolmer tests were those described in "A Manual of Serological Tests for Syphilis" 1949. Supplement No. 22, J. vener. Dis. Inform." For the V.D.R.L. and Eagle tests doubling dilutions to a titre of 1:256 of the test sera were used. For the Kolmer test the exact half-volume quantitative test was used with overnight fixation in the cold, employing serum dilutions as prescribed.

For the Price test, the original technique described by Price (1948) was used, with the exception that volumes were measured by pipette instead of using a dropper. The validity of the test is unaffected by this, as Price states that proportions and not actual volumes are important.

We used 0·02 ml. antigen and 0·1 ml. of each serum dilution of a titre of 1:256.

Antigens.—A "Wellcome" brand of cardiolipin antigen was used for the VDRL test. Eagle antigen was prepared locally and checked for reactivity by parallel titration with a single reference batch of Difco-Bacto antigen. Kolmer "improved" antigen was used as Cardiolipin was not available. It was prepared locally and checked against a single Difco-Bacto antigen. Price antigen was prepared locally and checked against a reference antigen kindly supplied by Dr. J. N. Orpwood Price.

Complement.—For the Kolmer test complement was prepared from the pooled heart blood of at least 100 guinea-pigs, diluted with an equal volume of boric acid-sodium acetate preservative and stored in 25-ml volumes at −20°C. It was diluted with 10 per cent. egg albumen in saline before use.

Controls.—A serum of known reactivity was freeze-dried in 0·5ml. quantities. A fresh ampoule was reconstituted as required. Whenever a batch of tests was performed, control titrations were made with a fresh ampoule of this control serum and also a non-reactive serum from the preceding batch of tests.

Recording of Results.—Results were recorded in dilutions (Harris, 1947), the number of dilutions being the reciprocal of the highest serum dilution found to be reactive. For the purposes of simplification, Price units (i.e. dilutions multiplied by five) were not employed. In the Kolmer test any degree of reactivity in the first tube was recorded as one dilution, in the second tube as two dilutions, in the third tube as four dilutions, in the fourth tube as eight dilutions, and in the fifth tube as forty dilutions.

Anticomplementary Sera.—Only three sera tested were anticomplementary. They were negative when retested using the modified Sachs method and did not react with any of the other three tests.

Results and Discussion

The sera which reacted were arranged in descending order of reactivity to the VDRL test, and numbered serially. The results are illustrated in the Figure (opposite). Each vertical column represents one
serum, and the height of the blocked column opposite the test represents the strength of the quantitative reaction to that test expressed in dilutions; 32 dilutions and over was considered as strongly reactive and was represented by blocking of the entire column. Four to 16 dilutions was considered to be moderately reactive and is represented by blocking half the height of the column. One to two dilutions is weakly reactive and is represented by blocking one quarter of the height of the column. The three smaller columns in the Kolmer series represent three doubtful reactions. If the serum was not reactive the column is unshaded. The Table summarizes the results. Of the 1,029 sera tested, 151 or 14.7 per cent. reacted to one or more of the four tests. There were 122 reactors to the VDRL test; 124 reactors to the Eagle test; 75 reactors to the Price test; and 87 reactors to the Kolmer test. Of the reactors, 57 or only 37.7 per cent. reacted to all four tests, whereas 38 or 25.2 per cent reacted to only one of the four tests. It can also be seen that by far the greater number of sera which reacted to one test only, were detected by the flocculation tests.

It is obvious from the Figure and Table that there is very little agreement amongst the results of the four tests. Even though there may be considerable numerical agreement, e.g. between the VDRL and Eagle tests, there is a great variation in the reaction of individual sera to these tests. Thus twenty reactors to the Eagle test (16.1 per cent.) failed to react to the VDRL, and eighteen VDRL reactors (14.8 per cent.) failed to react to the Eagle test. In this series of 1,029 individuals, 142 reacted with either the VDRL or the Eagle test or both. But using either test alone, only 104 of these would be detected with certainty irrespective of which test was used. Which of the remaining 38 reactors would be detected would depend entirely on the test used. Although the numerical correlation between these two tests is apparently good, namely, each detects about 120 reactors, the individuals concerned are different. Similarly, if the Price test were the test of choice, only 75 reactors would be detected, yet seven of them would have been acquitted by any of the other three tests. Again, if the 122 VDRL reactors are re-tested with the Kolmer test, 38 (31 per cent.) of them would fail to react.

Depending on which test or tests were used the estimated incidence of syphilis in this group would range from 7.5 to 14.7 per cent. Results published elsewhere indicate that the true figure probably lies between those two figures (Hill, Griffiths, and Buckle, 1957). Evaluated elsewhere (Kolmer, 1949; Price, 1949), the performance of each of these tests has been uniformly good, but in this series the results have been very irregular. Minor differences in techniques may cause great variation in serological results, but this is not likely to be the cause of these
discrepancies, as not only were the accepted techniques rigidly adhered to, but the most meticulous care was taken with each test. Freeze-dried serum of known reactivity was included as a control in each daily batch of tests. If any test was positive, all four tests were repeated, and throughout the trial there was no difference between the results of the two testings. All the tests were performed by specially-trained technicians of long experience.

If we assume that the specificity of each of these tests is as high in our series as in series overseas, then one must postulate that there is a high incidence of syphilis in the adult Bantu. This is not confirmed by autopsy figures (Hill and others, 1957). Also it would imply that there is a falling off in the sensitivity of these tests which failed to detect many of the reactors to others tests. As the antigens were known to be of a high standard, this decrease in sensitivity would have to be attributed to the presence of inhibitory factors in these Bantu sera.

Alternatively a decrease in the specificity of these tests could be postulated and this would appear to be the more reasonable hypothesis. But whatever the cause of these discrepancies, we cannot accept all these tests at their face value as aids to the diagnosis of syphilis.

A serum which reacts strongly to one test will not necessarily react strongly to all other tests. The Figure shows that, of the 22 sera which reacted strongly to one or more of the tests, only four reacted strongly to all four tests although sixteen others reacted to a lesser degree. Two of the sera were negative to one test whilst reacting strongly or moderately strongly to other tests. The greatest discrepancies occurred amongst sera which were reactive only in low dilution. Of the 38 sera which reacted to only one test, 27 did so at one dilution only. A large number of these reactions may perhaps have been due to non-specific factors widely distributed in Bantu sera.

In general, the correlation between the tests improves with the degree of reactivity of the serum. Concordance of positive results has been stated to be important in the diagnosis of syphilis (Targowsky, 1952) but this would seem to depend largely on the titre of the reactive serum. It is often assumed that the higher the titre of the reactive serum the more likely it is to be specific—this has found expression in the practice of “raising the diagnostic titre” (Shapera, 1955) of the test in order to improve its performance. This would also apparently improve the correlation between these tests. This assumption, however, is not necessarily correct, as there is no reason to think that a common reacting factor quite unrelated to syphilis would not on occasion be able to evoke a high antibody response. Also it must be remembered that the serologists have never suggested that a particular titre is necessary before a result is considered positive, as presumably their tests have been designed to conform to accepted standards of sensitivity and specificity.

On the basis of data collected during the course of this investigation, it was suggested that sensitivity and specificity ratings could be given for each level of reactivity for the four tests (Hill, Buckle, and Thomas, 1957). Certainly it is well known that many of the known sources of false positive reactions give only low titres of reactivity, and possibly they would be excluded at the higher levels. The figures quoted merely support this. But at the higher levels of reactivity the numbers of reactors are too small to be significant and it is probably unjustified on that evidence to claim a 100 per cent. specificity at a level of 16 dilutions, thereby implying that no non-specific factor is capable of inducing a strong antibody response.

Increase of specificity can only be achieved at the expense of missing many weak reactors, some at least of which may be specific, thus decreasing the value of the test to the clinician, although from the public health administrator’s point of view this point would be unimportant as weak reactors are likely to be less infectious. From the Table and Figure it can readily be calculated that no one of these four tests detected more than 82·1 per cent. of the sera which reacted to one or more of the other three tests. This excludes the use of any of them as a screen test.

The wide discrepancies in the results of standard tests when applied to the South African Bantu throws doubt on the reliability of them all when used in this racial group. Whether a serum is labelled “positive” or “negative” is of grave importance to the patient, yet what the serum is labelled appears to depend largely on which test is used. It would also seem that a single test used by itself, is not worth the time spent on doing it as it yields information of little clinical value. Possibly, in areas where there is a high incidence of clinical syphilis, and the true positives greatly out-number the false positives, it is justifiable to regard any reaction as significant, but in South Africa we have little knowledge of the incidence of such true and false positives and it is doubtful if such a serological result materially aids accurate diagnosis. Unfortunately, in many routine laboratories, technical errors may increase the number of false positives reported. Recently there has been a remarkable decrease in the incidence of clinical syphilis without, presumably, a decrease in the incidence of false reactions.
The introduction of antigens from a more specific source will not necessarily improve the performance of a test. Identical antigens can frequently be extracted from very different sources, and human sera, under certain circumstances, will react with antigens they have never previously encountered. Racial and geographic factors are probably further complicating factors, as it is well recognized that different racial groups have different serum protein patterns, and in different geographic areas such groups are exposed to very different antigenic stimuli. As these sera have been shown to act irregularly with some antigens, it is probable that they will react irregularly with all antigens. It is therefore unwise to assume without experiment that a serological test which correlates well with clinical findings in one population will do equally well in another type of population in another area.

It is often the practice of clinical laboratories, when changing or modifying their serological tests, to compare the results of the new test with the result obtained by the old one. Our results indicate that the rating of the new test will depend entirely on which test was in common use at the time of the comparison. Only if the results of one test are identical with those of another test over a big enough series, is it permissible to assume that the two tests have the same sensitivity and specificity. If they differ at all, it is impossible to assess whether one test is performing better than the other, unless we have definite knowledge of the true status, positive or negative, of each serum. Without this knowledge, by the time there have been three or four tests, each evaluated against the one before, it is impossible to have any idea of how the latest test compares with reality, even if the original test had been evaluated against clinical material—a precaution many laboratories omit. The authors therefore would urge all laboratories undertaking serological examinations to evaluate their results against clinical material as this is the only method of establishing the value of a test. This applies equally to the evaluation of any modification that might be introduced.

Conclusion

These results show that there are wide differences between the results of standard tests when used on South African Bantu sera. These differences appear to be greater than those observed when the same tests are applied to other racial groups in Europe or America, and indicate that the sensitivity-specificity ratings given to these tests on trials conducted elsewhere cannot be accepted here. Serology is an indirect method of detecting exposure to an antigen and, in view of the wide distribution in nature of similar antigens, it would be unwise to apply the results obtained in one population to another racial group which might have been exposed to a very different variety of antigenic stimuli. Differences in the ability of different racial groups to synthesize serum proteins for nutritional or other reasons may affect the serological behaviour. These results also show the fallacy of attempting to estimate the efficacy of a test by a numerical comparison with another test without a clinical trial. The discrepancies amongst these tests are so great that they render the tests unreliable for diagnosis. None of these tests can be assumed to be superior to any other test. None can be used as a screen test.

Summary

Sera from 1,029 apparently healthy adult South African Bantu males were subjected to four standard tests for syphilis. The results show a high incidence of reactors and a large degree of disagreement between the results obtained with the different antigens. The large number of reactors is unlikely to be due to a high incidence of the disease, but is more likely to be due to non-specific factors. It is suggested that different racial groups, because of their different antigenic experiences, would be expected to react differently to the same serological test.

The results illustrate the fallacy of comparing one test with another numerically without a clinical trial. Each test should be evaluated in the racial group in which it is to be used, as sensitivity-specificity ratings applicable to one racial group do not necessarily apply to another.

None of these tests is suitable for a screen test.

The authors wish to thank Dr. R. S. Dewar, venereologist to the King Edward VIII Hospital, Durban, for his advice and co-operation and also for the loan of a clinical assistant. We should also like to thank Dr. I. N. Orpwood Price for a sample of his antigen; Dr. G. D. English, Medical Officer of Health of Durban, and Dr. D. McKenzie for allowing us to collect the blood samples; the Provincial Pathologist, who allowed us to use the facilities of the Provincial Laboratories for the investigation, and the Director of Medical Services for permission to publish the results.

REFERENCES