REVALUATION OF THE FLOURPAPER TUBERCULIN JELLY TEST

BY

MAXWELL CAPLIN, M.R.C.S., L.R.C.P.

Tuberculosis Officer, London Chest Hospital

JUNE HARRINGTON, M.B., B.S.

C. P. SILVER, B.M., M.R.C.P.

Assistant Tuberculosis Officers, London Chest Hospital

AND

STEFAN GRZYBOWSKI, M.D., M.R.C.P.

Senior Registrar, High Wood Hospital for Children

The tuberculin jelly test is widely used, and with the introduction of B.C.G. its use has been further extended. We in clinic practice, however, have come to distrust the results of the jelly test, and there has been the same dissatisfaction at High Wood Hospital for Children. No fewer than 17 out of 200 children referred there recently for treatment for primary pulmonary tuberculosis proved to be non-tuberculous and tuberculin-negative; the erroneous diagnosis in each of these cases had been based on a jelly test read as positive. We decided, therefore, to carry out an investigation comparing the accuracy of the jelly test with that of the Mantoux.

, History

The use of tuberculin jelly was described by Monrad in Denmark in 1936, and the whole early history of the test was first discussed in this country in a leading article in Tubercle (1937). Paterson (1944), using a jelly containing 95% old tuberculin (O.T.), considered it equivalent to a Mantoux test of 1:100 dilution. This degree of sensitivity was not generally accepted, and in an effort to increase the sensitivity of the test Deane (1946) tried sandpapering the skin lightly before the application of the jelly. paper was later replaced by flourpaper (gauge 00 sandpaper). This technique is generally used in children over 5, while for those under 5 the plain jelly is considered quite adequate (Dick, 1950a). Dick (1950b) comparing P.P.D. and O.T. jellies of various strengths, found the 60% O.T. jelly Lendrum (1951) has claimed great the more reliable. accuracy in reading the results, particularly if the test be read as late as the fifth day. In contrast, Coles (1952) and Frew, Davidson, and Reid (1953) have reported a high rate of doubtful and inaccurate results.

One of the major difficulties in reading the test, particularly the flourpaper method, has been that of deciding what skin appearances are essential for a positive result, for each authority has laid down different standards. The main standards that have been advocated are given in Table I.

The present investigation consisted of three parts: (1) a detailed study of a large number of plain and flour-paper jelly tests with simultaneous Mantoux tests; (2) the independent reading of a series of flourpaper jelly tests and of Mantoux tests by a panel of observers who had no special experience of the tuberculin tests; and (3) the independent reading of flourpaper jelly tests and of Mantoux tests by a panel of expert observers.

The Jelly Test

Plain and flourpaper jelly tests and Mantoux tests were performed together on 288 children (ages 5 to 15) from two East London schools and from High Wood Hospital for Children. The schools provided a predominantly negative population, and the hospital provided children with active clinical tuberculosis. Of these 288 children, 173 were negative to a 1:100 dilution of O.T., and 115 were positive to 1:100 or higher dilutions of O.T. (Table IV). No fewer than 92 of this latter group were cases of active tuberculosis undergoing treatment in High Wood Hospital. As there is a tendency for active tuberculosis to be associated with high levels of tuberculin sensitivity (Grzybowski, 1951), the Mantoux-positive group showed on the whole a much higher sensitivity than would be the case were it derived wholly from the general population.

The jelly used contained 60% O.T. and was manufactured by Allen and Hanburys. All the jelly tests were applied by one observer in the manner described in the makers' instructions. The interscapular area was cleaned with acetone, and control jelly was squeezed out in a horizontal line about $\frac{3}{4}$ in. (2 cm.) long to the left of the spine. Below the control, tuberculin jelly was applied in the form of a V, the arms of which were also about $\frac{3}{4}$ in. (2 cm.) long. On the right side of the spine the whole area was lightly sandpapered by six strokes with No. 00 sandpaper (flourpaper). and the control and the tuberculin jelly were applied in the same way. Zinc oxide strapping was placed directly over the jelly, and removed on the third day. The reaction was examined on the third, fourth, and fifth days and recorded in a standard manner, the results being read together by at least three of us. In each test the presence of erythema and of oedema (induration, raising) was recorded; these features concerned the base of the lesion. The elevations upon the base were recorded as papules, vesicles, vesicopapules, and pustules. Staining, crusts, and scratch marks were also noted.

The Mantoux Test

The tuberculin dilutions were freshly prepared from old tuberculin (B.W. Batch No. T.2732.S), and separate syringes and needles were used for each dilution. To determine the degree of tuberculin sensitivity the following procedure was adopted. In the predominantly negative population of the schools an intradermal injection of 0.1 ml. of 1:1,000 dilution (10 T.U.) was given on the first day at the time of the jelly testing. The result was read on the third day (48 hours); if negative, the test was repeated with 1:100 dilution (100 T.U.); if positive, with both 1:10,000 and 1:100,000 dilutions (1 T.U. and 0.1 T.U. respectively). The

TABLE I.-Varying Standards for Judging a Positive Jelly Reaction

Author	Method	Time of Reading Recommended	Standard for Positive Result	Remarks
Paterson (1944)	Plain jelly	48 hours and several days thereafter	Erythema or slight vesiculation in shape of marking used	Negative result showed a completely
Clark (1951)	, ,,	72 or 96 hours	Definite erythema with at least two vesicles; erythematous area often palpably indurated	
Deane (1946)	Sandpaper jelly	96 hours	None	No. 0 sandpaper used. All late observers have used No. 00
National Tuberculin Survey 1949-50 (M.R.C., 1952)	Flourpaper jelly	72 to 96 hours	Raised erythema with or without vesiculation; coalescence of vesicles to form V- or heart- shaped area of induration might occur	
Dick (1950b)	,,	72 to 96	Vesiculation. Coalescence of vesicles to form V- or heart-shaped area of induration might occur	Fewer than four vesicles should be considered a doubtful reaction Erythema should be ignored
Lendrum (1951)	,, ,,	96 hours	Six or more papules in shape of original application	Erythema should be ignored
Makers' instructions (Allen & Hanburys)	Plain or flour- paper jelly	72 hours and there- after up to one week	Slightly vesicular mark	Erythema should be ignored

results were read on the fifth day (48 hours later). The children from High Wood were tested with 1:10,000 and 1:100,000 dilutions (1 T.U. and 0.1 T.U. respectively) on the first day, and the results were read on the third day (48 hours). Only one High Wood child was negative to the 1:10,000 dilution (1 T.U.) and he was retested with the 1:1,000 dilution (10 T.U.) and shown to be positive.

A reaction two to four days after injection, with not less than 5 mm. diameter of oedema with surrounding erythema, was regarded as a positive Mantoux result.

Development of the Skin Lesion

The development of the plain jelly lesion can be used to explain that of the more complicated flourpaper reaction. In the Mantoux-positive group, if the reaction to the plain jelly was severe, the area covered by the jelly was raised, reddened, and oedematous. This oedema was most pronounced at 72 hours. Vesicles were appearing at 48 hours, but tended to be small, increasing in size and number by 96 hours. Papules were much less common, and tended to be more numerous initially. Pustules were rare; crusts were never seen. When the reaction was less severe there were separate islands of reddening and oedema, often surmounted by one or more vesicles. The incidence of each type of skin change is given in Table II, which shows that

TABLE II.—Percentage Distribution and Type of Plain Jelly Lesion According to Mantoux State (1:100)

Reaction	Mantou 1	x-positiv 15 Chiidr	e Group en	Mantoux-negative Group 173 Children				
Reaction	48	Hours 72	96	48	Hours 72	96		
Erythema Oedema Staining	63 27	69 48	72 48	<u> </u>	_	_		
Vesicles Vesicopapules	56 17	66 10	77 3	=	=	=		
Papules Pustules Crusts	14 1	3	3 1		_	=		
Scratch marks	_	_	=	=	_			

a reaction of one type or another occurred in the majority of the Mantoux-positive children, while there was no reaction at all in the Mantoux-negative group at 72 and 96 hours. There is little doubt that this high degree of correlation in our Mantoux-positive group is due to the high levels of sensitivity in this group; with the lower levels of sensitivity encountered in the general population more cases would show a negative reaction. It is for this reason that the use of flourpaper was originally introduced.

When the flourpaper technique was used the reactions in the Mantoux-positive group were more violent than those seen with the plain jelly. At 48 hours there was intense erythema and oedema, with pustules rather than vesicles. By 72 hours the whole reaction was more pleomorphic than that of the plain jelly. The surface was oozing serum, and many vesicles and pustules had burst. At 96 hours the process had gone further, and often the base was flat though covered by dried or drying vesicles or by crust. If the jelly response was less marked the reaction was patchy, and in many cases was hard to differentiate from

Table III.—Percentage Distribution and Type of Flourpaper Jelly Lesion According to Mantoux State (1:100)

Reaction	Mantoux 11	-positive 5 Childr	e Group en	Mantoux-negative Group 173 Children			
Reaction	Hours 72		96			ours 72 96	
Erythema Oedema Staining Vesicles Vesicopapules Papules Pustules Crusts	95 79 — 22 6 3 70 5	90 69 1 31 5 3 30 57	84 41 10 24 4 1 4 63	68 9 1 11 11 13 48 12	34 2 4 1 - 9 2 25 53	11 25 — 1 39	

that seen in the Mantoux-negative cases. Despite careful use of the flourpaper the majority (all but 28) of the 173 Mantoux-negative children showed some reaction at some time (see Table III). In these Mantoux-negative children a reaction elicited by the control jelly was much less often seen and was always less severe than the reaction due to tuberculin jelly, though one would have expected them to be the same. We are unable to account for this surprising finding, as the only apparent difference between tuberculin and control jelly is that of tuberculin content. It may be that the position in which the two jellies are placed in relation to the force and direction of the stroke of the flourpaper may provide a clue. The fact remains, however, that in this series the control jelly was of no value.

If Tables II and III are compared it is apparent that, whilst the plain jelly lesion is relatively simple, the flour-paper lesion is a very complex one. The frequency of pustules at 48 hours and the commonness of crusting and scratch reactions in both positive and negative reactors form a sharp contrast with the plain jelly reaction. All the lesions lying in the Mantoux-negative half of Table III can only be due to the use of flourpaper.

Various Standards Applied to the Skin Lesion

Our findings may be compared with the standards laid down by different authorities. Two commonly accepted standards have been chosen, and are named Standard A and Standard B.

Standard A.—A positive reaction: an area of raised erythema (erythema and oedema) with or without vesiculation. A negative reaction: any reaction other than this.

This was the standard accepted by the Medical Research Council (1952) in the National Survey 1949-50.

Standard B.—A positive reaction: at 72 hours four or more skin elements—that is, vesicles, vesicopapules, papules, and pustules—and at 96 hours six or more skin elements; or an area of raised erythema at either 72 or 96 hours considered to be due to coalescence of vesicles. (The substitution of six elements at 96 hours for four at 72 hours affected only three cases.) This standard is similar to the standards laid down by Dick (four or more vesicles) and Lendrum (six or more papules) (see Table I). There seems little doubt that the vesicles of Dick are the papules of Lendrum.

Table IV shows the results when Standard A and Standard B are read at 72 and 96 hours.

Table IV.—Flourpaper Jelly Results. Comparison with Mantoux Sensitivity

	No. of	72 H	lours ·	96 Hours			
Mantoux Sensitivity	Children	Positive to Standard A	Positive to Standard B	Positive to Standard A	Positive to Standard B		
1: 100,000 + 1: 10,000 + 1: 1,000 + 1: 100 +	64 39 7 5	47 (73%) 27 (69%) 2 (29%) 0 (0%)	62 (97%) 32 (82%) 3 (43%) 0 (0%)	27 (42%) 13 (33%) 2 (29%) 0 (0%)	38 (59%) 20 (51%) 3 (43%) 0 (0%)		
Positive	115	76 (66%)	97 (84%)	42 (37%)	61 (53%)		
Negative	173	0 (0%)	18 (10%)	0 (0%)	0 (0%)		

As might be expected, the proportion of children showing a positive reaction to the jelly is closely related to the level of tuberculin sensitivity. But even in those showing a high sensitivity the jelly test was sometimes negative. Thus it does not appear permissible to state that the jelly test is equivalent to a particular dilution of tuberculin. In considering Table IV it must be remembered that this group of Mantoux-positive children is selected. Were these children a representative sample of the population the level of sensitivity would be lower and the discrepancy between Mantoux and jelly tests among the tuberculin-positive children would be greater.

Table IV shows that by applying these two standards at 72 hours there will be a considerable error in reading the test; using Standard A, 34% of tuberculin-positive children will show a negative reaction, but there will be no false-positive reactions among the tuberculin-negative children.

The introduction of Standard B will diminish the number of false-negative reactions to about 16%, but will lead to the occurrence of about 10% of false-positive readings. At 96 hours, using either standard, the number of false positives falls to zero, but the number of true positives also falls—that is, the false negatives increase to 63% according to Standard A and 47% according to Standard B. Applying other standards, none was found to give satisfactory results. It was easy to shift the body of the readings from negative to positive or from positive to negative by accepting or rejecting some particular factor.

The Non-expert Panels

Is the flourpaper jelly test as easy to read as its advocates maintain? This part of the investigation was designed to answer this question by showing what accuracy and correlation could be achieved by relatively inexperienced observers.

It was arranged that at one of the schools a panel of four doctors should read the flourpaper results and the Mantoux results on the fourth day (72 hours), and another panel of similar constitution should read both tests on the fifth day (96 hours). Each panel consisted of three general practitioners and one paediatric registrar. Each doctor was provided with the makers' instructions for reading the flourpaper jelly test and also instructions for reading the Man-They read the tests independently, recording toux test. their findings as positive, negative, or doubtful. All the flourpaper jelly tests were read first, only the children's backs being exposed, and then all the Mantoux tests, only the left forearm being exposed. Of the 108 children examined, 101 were negative to a 1:100 dilution of O.T., and 7 were Mantoux positive.

Table V shows that there was little disagreement between the four doctors of each panel when reading the Mantoux test, but their disagreement over the jelly readings was very

TABLE V.—Percentage Agreement Between Non-expert Observers

	Reading	at 72 Hours	Reading at 96 Hours		
•	Jelly	Mantoux	Jelly	Mantoux	
Unanimous agreement 3 out of 4 doctors agree Less than 3 agreements	 31% 34% 35%	78% 18% 4%	26% 42% 32%	96% 2% 2%	

considerable. Unanimous agreement was reached in less than a third of the jelly readings. If three agreements are accepted as a satisfactory degree of correlation, 68% agreement is all that was achieved by the more consistent panel—that is, the 96-hour panel—compared with 98% agreement by the same panel when reading the Mantoux test.

Turning now to the individual readings of each observer, Table VI gives their performance related to the Mantoux state as determined by us.

Table VI.—Non-expert Panels. Individual Performance of Each Observer Compared with the Mantoux State (1:100)

	101 Mantoux-negative Children						7	7 Mantoux-positive Children								
	Jelly Re							ead	ing a	at						
		72 Hours 96 Hours					7	2 H	our	s		96 I	Iour	s		
Observer:	A	В	C	D	E	F	G	н	A	В	С	D	E	F	G	Н
False ,, Readings re- corded as	85 8	75 6	34	33	63	72 26	84	39 60	5	5 2	4 3	1 0	6	1	6	6
doubtful	8	20	11	28	3	3	17	2	1	0	0	6	0	5	0	0

Only three out of the eight observers from both panels secured a fair correlation between their jelly readings and the Mantoux state. Both registrars (Dr. A and Dr. G), if their doubtfuls be disregarded, secured some impression of the tuberculin state, but even the more accurate of the two

recorded as many as 17 doubtfuls and the less accurate 8 false positives. One practitioner (Dr. B) approached their performance, but the remaining practitioners showed remarkable variations in their readings, two reading 60 or more false positives and one being in doubt on 34 occasions. The lack of correlation shown in Table V is not, therefore, a matter of one observer disagreeing with all the others, but is due to the variety of opinions expressed about each result.

In contrast, reading the Mantoux tests the panels achieved good agreement, not only among themselves but also with our readings. Three out of the eight completely agreed with our readings, and three disagreed only twice; the largest number of times any observer disagreed was 21 times, but he recorded 20 results as doubtful.

The Expert Panels

This part of the investigation was designed to answer two important questions: (1) Could two experts in the use of the flourpaper jelly test obtain a high degree of correlation between their readings? (2) What correlation is there between the flourpaper jelly test and the Mantoux test when these tests are done at the same time, on the same individual, the tests being performed and read by doctors experienced in each?

To undertake this investigation we enlisted the aid of a chest physician who had written favourably about the flour-paper test. She asked a colleague with similar experience to join her in forming the "jelly panel." Two of us acted as the "Mantoux panel."

The school chosen for the investigation was a girls' grammar school in East London with 450 pupils, whose ages ranged from 10 to 17: 437 of the girls took part in the investigation.

Each girl had a flourpaper jelly test applied by one member of the jelly panel, and received a Mantoux test (1:1,000) given by one member of the Mantoux panel.

On the third day (48 hours later) the Mantoux tests were read independently by the two doctors of the Mantoux panel. They were seated at opposite sides of the examination room and were separated by a screen. The pupils filed past each in turn, exposing only the left forearm, and the results were recorded on specially prepared lists. These lists, when completed, were handed to the schoolmistress in charge of the arrangements; she checked the correlation at the end of the session.

On the fifth day (96 hours) the jelly panel, taking the same precautions, read their jelly tests. This time the pupils' forearms were covered and only the backs were exposed.

Table VII illustrates the correlation between the Mantoux observers and that between the jelly observers.

Table VII.—Showing Correlation Between Mantoux Observers and Between Jelly Observers

No. of Children		on between Observers		on between bservers
Tested	Agreement	Disagreement	Agreement	Disagreement
437	436 (99.8%)	1 (0.2%)	277 (63·4%)	160 (36.6%)

It will be seen that the Mantoux panel secured 99.8% agreement, whereas the jelly panel secured only 63.4% agreement.

The correlation between each jelly observer and the Mantoux state (1:1,000) can now be considered. Of the 437 children tested the Mantoux results of two were omitted; in one, both Mantoux observers recorded their only doubtful result; in the other, the Mantoux observers recorded their only disagreement. The remainder consisted of 435 children whose Mantoux state had been determined and agreed upon. The performance of each member of the jelly panel was then compared with this Mantoux state.

898 Ост. 16, 1954

Table VIII illustrates the poor correlation that was obtained. Dr. X recorded no fewer than 123 (34%) of the Mantoux-negative children as positive, and was in doubt

TABLE VIII.—Showing Correlation Between Each Jelly Observer and the Mantoux State (1:1,000)

Mantoux State	Jelly	Observer :	Dr. X	Jelly Observer Dr. Y			
(1: 1,000 O.T.)	Positive	Negative	Doubtful	Positive	Negative	Doubtful	
74 positive children 361 negative child-	68	4	2	61	11	2	
ren	123	185	53	32	325	4	

on 55 occasions; however, she recorded only four (5%) false-negative readings. Dr. Y gave 32 (9%) false-positive readings, but 11 (15%) of the Mantoux-positive children were recorded as negative. Subsequently Dr. X, the observer who had applied the jelly tests, attributed the unsatisfactory results to hurried application of the tests and over-scarification of the skin. Both Mantoux and jelly tests were completed in just over three hours, as the working conditions were excellent and no delays occurred. In our opinion hurried application could not account for the discrepancy between the readings of the two jelly observers.

Discussion

The first part of this investigation traced the development of the skin reactions resulting from the plain and the flourpaper jelly tests applied in the same individual simultaneously. It illustrated the range and type of the lesions produced, and it enabled us to correlate the findings with the standards laid down by various authorities. taneous Mantoux tests with graded dilutions of O.T. allowed the comparison between the two tests at various grades of sensitivity. Our findings showed that the flour paper jelly produced a complex reaction consisting of numerous elements of which none was specific to the Mantoux-positive members of our series. No single factor or group of factors could be obtained which would coincide with the Mantoux findings. Whichever standard was employed produced its quota of false positives or false negatives or both, and any attempt to reduce the one increased the other. Nor was control jelly of any real value in distinguishing the false from the true reaction.

The second part of the investigation illustrated the widely differing results obtained when the flourpaper jelly test is read by those without special experience of the test. The non-expert panel were unanimously agreed on their findings in fewer than one-third of the tests at 72 hours, and at 96 hours another panel achieved unanimous agreement in only one-quarter.

The third part of the investigation demonstrated clearly that the jelly test, even in expert hands, cannot be read accurately. The jelly panel, consisting of two experts reading the test which one of them had performed, disagreed 160 times out of 437 (37%). This is a serious lack of agreement. It is obvious that in expert hands an unsatisfactory correlation can only mean an unsatisfactory test.

These results show that the flourpaper jelly test is not accurate enough for testing either contacts or those suspected of tuberculosis. A false-positive reading is common and will result in a child contact not receiving B.C.G. vaccination. Again, if such a child develops some minor pulmonary infection associated with radiological changes the mistaken diagnosis of tuberculosis may be made. A false-negative result also carries its hazards, for in many clinics and hospitals such a result is not always followed by an intradermal Mantoux test, and the child may not be considered to require a chest x-ray examination in the absence of symptoms. A case of tuberculosis may thus remain undiagnosed.

The use of the jelly test in surveys also has little to commend it. It is obvious that, for accurate results to be obtained, the standard of a positive reaction must be strict in order not to allow any false-positive readings among the tuberculin-negative children. In such circumstances many tuberculin-positive children will be missed, and it is therefore essential to perform an intradermal test subsequently. When only a minority of the children in a community have been infected with tuberculosis, as is the case in this country, then in any group of children under survey the proportion found to be tuberculin-positive through the jelly test will be small. In other words, the sensitivity of the ielly test when read according to rigid standards to avoid the false-positive error is too low to make it a satisfactory screening test, because the great majority will, anyway, require a subsequent Mantoux test.

Certain surveys have been carried out with the jelly test alone, without subsequent Mantoux tests. If the standards of the positive result were similar to those applied by the observers in the second and third parts of this investigation, both false-positive and false-negative readings would be obtained. They would tend to cancel each other to a varying degree, depending on the standards adopted and on the proportion of tuberculin-positive children in the population. In certain circumstances the end-result of such investigations could give a reasonable figure for the tuberculin state of the population concerned, but this is obviously an unsatisfactory way of conducting a tuberculin survey.

Tuberculin jelly is cheap and stable, it can be easily applied, and it possesses the great advantage of being painless, but all these virtues are of no account if the test cannot be read accurately. The inaccuracy of the flourpaper jelly test readings in both expert and non-expert hands precludes its use both in clinic practice and in tuberculin surveys.

Summary

Plain and flourpaper jelly tests were performed on 288 children aged 5 to 15 whose tuberculin sensitivity had been determined by Mantoux tests with graded dilutions of O.T. The jelly lesions were examined on three successive days. The flourpaper test findings were compared with various standards advocated for judging positive and negative results. Whichever standard was used produced false results.

Non-expert observers independently reading the flourpaper tests of 108 of these children at 72 and 96 hours did not agree among themselves or with the Mantoux state. They agreed well over their Mantoux results.

Flourpaper jelly and Mantoux tests (1:1,000) were performed simultaneously on 437 girls aged 10 to 17. Two expert observers reading the jelly tests independently disagreed 160 times. Two other observers reading the Mantoux tests independently disagreed only once.

It is concluded that the flourpaper jelly test is unsatisfactory both in clinical practice and in tuberculin surveys.

We thank the members of both panels and the headmistresses of the schools for their generous help; Dr. F. J. Bentley for permission to carry out the investigation at High Wood; Dr. K. F. W. Hinson and Dr. N. Lloyd Rusby for advice; Mr. B. Benjamin for advice and statistical aid; and Miss E. Hewison, Sister B. E. Hill, and the staffs of the Tuberculosis Dispensary, London Chest Hospital, and of High Wood Hospital for their

REFERENCES

Clark, N. S. (1951). Lancet, 2, 464. Coles, H. M. T. (1952). Ibid., 2, 927. Deane, E. H. W. (1946). Ibid., 1, 162. Dick, W. P. (1950a). British Medical Journal, 2, 1333. —— (1950b). Ibid., 2, 141.

Frew, H. W. O., Davidson, J. R., and Reid, J. T. W. (1953). Brit. J. Tuberc., 47, 68.
Grzybowski, S. (1951). Tubercle, Lond., 32, 145.
Lendrum, J. D. (1951). British Medical Journal, 2, 148. Medical Research Council (1952). Lancet, 1, 775. Monrad (1936). Ugesk. læg. 98, 885. Paterson, D. (1944). British Medical Journal, 1, 531. Tubercle., Lond., 1937, 18, 172.