Bull. Org. mond. Santé } 1971, **45**, 543-550
Bull. Wld Hlth Org. }

A Co-operative Evaluation of Test Systems Used to Assay Tuberculosis Vaccines[1]

E. H. WIEGESHAUS, G. HARDING, D. McMURRAY, A. A. GROVER, & D. W. SMITH

On Behalf of the Participating Laboratories

The results of research on the immunogenicity of experimental mycobacterial vaccines are characterized by a remarkable lack of agreement about which substances are most immunogenic. The disagreement has usually been attributed to the differences in the methods of preparing the vaccines. An alternative hypothesis is that the conflicting results are a product of the different methods used to assess the potency of the vaccines.

To determine if the method by which a vaccine is tested is a major factor contributing to the disagreement, an experiment was conducted in which a series of five different vaccines was distributed to each of nine participating laboratories. Each investigator evaluated the potency of the vaccines in one or more animal models of his own choosing. This in effect held the method of vaccine preparation constant while permitting all other variables to change.

The ranking of the five vaccines was random, thus demonstrating that the method by which a vaccine is tested influences the apparent potency of a vaccine. These results cast doubt on the conclusions about the relative potency of tuberculosis vaccines evaluated by different methods.

INTRODUCTION

Interest in the problem of how to measure the effectiveness of tuberculosis vaccines in the laboratory was renewed in part as a result of a meeting of WHO Advisers on Immunogenic Agents in Tuberculosis in Geneva in 1964. Consideration of a brief review of the literature on non-living vaccines indicated a lack of agreement on which substances are most immunogenic and suggested that variations in methods of measuring immunity may be a contributing factor. This group of advisers recommended the use of sensitive, reliable, and valid protection tests in animals and that the further development of such methods should receive considerable attention. A more comprehensive review (Smith, Grover & Wiegeshaus, 1968) also concluded that there was a lack of agreement on the immunogenicity of non-living vaccines and fractions of tubercle bacilli. The conclusion commonly drawn from an analysis of the conflicting results from the various laboratories testing immunogenicity of either living or non-living tuberculosis vaccines is that the lack of agreement stems from differences in the method of preparing the vaccines. An alternative conclusion is that the conflicting results are a product of different methods of assessing the potency of vaccines. Whether the variation in test methods is an important factor could be determined by comparing vaccines in several different animal test systems.[2] Usually, in assaying tuberculosis vaccines, a single test system is used to compare vaccines. In

[1] This paper reports a co-operative evaluation in which the institutions and persons listed in the Annex took part: the study was planned and carried out in consultation with the World Health Organization, Geneva. This report was prepared by the group at the University of Wisconsin. The contributions to this project by the University of Wisconsin group were supported in part by the United States-Japan Co-operative Medical Science Program (Contract Ph 43-67-739), the National Institutes of Health, Department of Health, Education, and Welfare, Bethesda, Maryland, and by funds from the Madison Tuberculosis Association, Madison, Wis., USA.

[2] A test system is defined to include the specific choices for the dependent and independent variables selected for the purpose of evaluating vaccines (the animal model) plus other variables associated with a particular laboratory (lab). Thus: test system = animal model + lab.

The independent variables of the animal model would include animal species, vaccination route, vaccination–challenge interval, challenge route, etc. The dependent variable(s), i.e., some measure of the response to infection, would include survival time, spleen index, the number of mycobacteria recovered from lung or spleen, etc.

contrast, in this study five vaccines were used to compare several test systems.

METHODS AND MATERIALS

The 9 co-operating laboratories each received the same vaccines to evaluate. Laboratories having experience with more than one animal model were encouraged to assay the vaccines in several animal models. In an attempt to reduce the number of possible variables without significantly altering any animal model, agreement was reached to assign a specific value to three of the possible animal-model variables—namely, vaccination dose, vaccination–challenge interval, and challenge strain.[1] This left the following variables open to the choice of each participating laboratory: animal species, vaccination route, challenge–sacrifice interval, challenge route, challenge dose, and response to infection to be measured.

Vaccines

The vaccines were prepared in the Biological Laboratories, Glaxo Laboratories, Ltd., Greenford, Middlesex, England, and were sent to the Tuberculosis Research Institute, Prague, Czechoslovakia, for distribution to the participants.

Vaccine A: live BCG lot F-10. This batch of vaccine was prepared by Glaxo and was purchased by the World Health Organization. Organisms of the "Copenhagen" strain were grown in deep culture and were used to prepare vaccine freeze-dried in dextran–glucose–Triton according to the method described by Ungar et al. (1962). After reconstitution the vials contained 0.75 mg/ml moist weight of vaccine with a viability count of 10.7×10^6/ml.

Vaccine B: live Mycobacterium avium. A subculture of the "Weybridge" strain PHS-262 of *Myco. avium* was obtained from Dr Carroll Palmer, Department of Health, Education, and Welfare, Atlanta, Ga., USA. It was the original intention to prepare a freeze-dried vaccine from this strain as was done for the live BCG vaccine. This proved impossible because the strain would not grow in the glycerol-free culture medium used for BCG. The strain was therefore grown on the surface of Löwenstein–Jensen medium and harvested after 5 weeks' incubation at 37°C. The cells were suspended by shaking with glass beads in dextran–glucose–Triton solution, which was then processed as indicated above. After reconstitution the vials contained 0.75 mg/ml moist weight of vaccine with a viability count of 3.0×10^6/ml.

Vaccine C: formalin-killed Mycobacterium tuberculosis, *saline suspension. Myco. tuberculosis*, strain H37Rv, was grown on the surface of liquid synthetic medium for 8 weeks. The cells were harvested by centrifugation, suspended in 50 ml of 1% formalin saline, and heated at 60°C for 30 min. The suspension was homogenized and left at 25°C overnight. The suspension was diluted with 0.15 M NaCl containing 0.013% merthiolate (merthiolate–saline) to a concentration of 5 mg/ml moist weight and dispensed in 5-ml amounts into rubber-capped vials. The vaccine was tested for sterility by culture and by inoculation into guinea-pigs. The preparation was shown to be free of viable *Myco. tuberculosis.*

Vaccine D: formalin-killed Myco. tuberculosis, *suspended in oil-Arlacel.* Part of the material that formed Vaccine C was diluted 1 : 3 in the merthiolate–saline solution. Three parts of aqueous vaccine were then emulsified with 10 parts of the oil-Arlacel mixture to give a final concentration of 0.5 mg/ml moist weight.

Vaccine E: "extraction residue" suspended in oil-Arlacel. The extraction residue was supplied as a dry powder by D. W. Smith, University of Wisconsin, Madison, Wis., USA. The preparation of the extraction residue was described by Fregnan & Smith (1963). A quantity of 1200 mg of the dry powder, suspended in 100 ml of merthiolate–saline, was homogenized and then further diluted with merthiolate–saline to 720 ml. This solution was then mixed with oil-Arlacel to give a final concentration of the extraction residue of 0.5 mg/ml.

Details of the preparation of the vaccines in adjuvant

The oil phase consisting of 10% w/v Arlacel A (Atlas Chemical Industries, Wilmington, Del., USA) in mineral oil (Puremore, Burmah Oil Trading, Ltd., London, England) was sterilized by filtration through a membrane filter of 0.22-μm pore size. The oil phase was transferred to a beaker and the aqueous phase containing the vaccine was allowed to run in gradually during constant vigorous stirring with a Vortex mixer. The vaccine was then homogenized by passing it twice through a Minisonic Homogenizer (Ultrasonics, Ltd.) and then distributed into vials in 6-ml amounts. The vials were capped with oil-resistant rubber caps and sealed.

[1] Some laboratories found it necessary to substitute a different challenge culture.

Reconstitution of the vaccines, dosages, and vaccination schedules

Vaccine A : live BCG-F10. Each ampoule after reconstitution with 5 ml of saline contained 0.075 mg/0.1 ml. A single dose of 0.1 ml was given 6 weeks before challenge.

Vaccine B : live Myco. avium. Each ampoule of freeze-dried vaccine after reconstitution with 5 ml of saline contained 0.075 mg/0.1 ml. A single dose of 0.1 ml was given 6 weeks before challenge.

Vaccine C : formalin-killed H37Rv suspended in saline. Each ampoule contained 5 ml of vaccine at a concentration of 5 mg/ml. Two doses of 0.1 ml (0.5 mg) were given 3 weeks apart. The second injection was given 3 weeks before challenge. If possible, an additional group of animals was given 2 doses of the same vaccine diluted 1 : 10 with saline (each animal received 0.05 mg at each injection).

Vaccine D : formalin-killed H37Rv suspended in adjuvant. Each ampoule contained 0.5 mg/ml of vaccine in adjuvant. Two doses of 0.1 ml (0.05 mg) each were given 3 weeks apart. The second injection was given 3 weeks before challenge.

Vaccine E : Extraction residue suspended in adjuvant. Each ampoule contained 0.5 mg/ml of vaccine in adjuvant and 2 injections of 0.1 ml were given 3 weeks apart. The second injection was given 3 weeks before challenge.

Preparation of challenge suspension

In order to standardize the virulence of the challenge culture, participants were requested to obtain a fresh transplant of *Myco. tuberculosis* strain H37Rv from the Trudeau Institute, Saranac Lake, N.Y., USA. The culture was maintained as a pellicle on Sauton medium and subcultured every 14 days. It was recommended that the challenge suspension be prepared from a culture grown for not more than 14 days on Sauton medium or from a subsurface culture grown for not more than 10 days in Dubos broth.

Test systems used to evaluate vaccine

Table 1 provides a summary of the important parameters of some of the various test systems that were compared.

Table 1. Comparison of the important parameters for 12 of the 21 animal test systems studied [a]

Test system	Animal species [b]	Vaccination route [c]	Challenge route [c]	Challenge dose per animal (viable units)	Challenge–sacrifice interval (weeks)	Response to infection measured
1	Red mice (50)	SC	IV	1.58×10^{4}	—	Survival
2	Mice (6)	IV, IP [d]	IV	6.0×10^{4}	3	Microbial enumeration
3	Mice (40)	IP	IV	0.5 mg	6	Lung score
4	Guinea-pigs (30)	IP	SC	1.0×10^{5}	4	Spleen index
5	Mice (25)	SC	IV	2.0×10^{6}	—	Survival
6	Mice (25)	SC	Resp.	30–40	4	Microbial enumeration
7	Mice (80)	IP	IV	1.0 mg	—	Survival
8	Mice (20)	SC	Resp.	5–10	4	Microbial enumeration
9	Guinea-pigs (20)	SC	Resp.	5–10	5	Microbial enumeration
10	Guinea-pigs (31)	SC	SC	1.6×10^{5}	6	Feldman Index
11	Rats (18)	IP	IV	1.3×10^{6}	—	Survival
12	Mice (43)	IP	IV	1.3×10^{6}	—	Survival

[a] The other test systems differed primarily in the response to infection measured. Some laboratories measured up to 4 different responses to infection.

[b] The number in parenthesis is the number of animals per treatment group.

[c] SC = subcutaneous; IV = intravenous; IP = intraperitoneal; Resp. = respiratory.

[d] Vaccines in adjuvant given by the intraperitoneal route.

Statistical analysis

The experiment was carried out according to a randomized complete block design. The treatment term in the analysis of variance is testing for differences between vaccines. The orthogonal individual degree of freedom technique (Li, 1964) was used to separate the 5 degrees of freedom of the treatment effect into 5 specific comparisons between vaccines. Because the participating laboratories measured different responses to infection, it was necessary to use the rank order of the vaccines to obtain a common basis of measurement in order to compare directly the results of different laboratories as well as the results obtained with different animal models within one laboratory. In order to conduct a valid parametric statistical analysis the data were transformed to standard normal scores. For the calculation of the treatment and interaction effects in the analysis, the means of the treatment groups were ranked in order of increasing severity of disease regardless of group and transformed to standard normal scores. To calculate a " within " term, the data from each animal, from those animal models that included a continuous dependent variable, were ranked in the order of increasing severity of disease regardless of group and then transformed into standard normal scores. These transformations were made using the tables published by Harter (1961).

Table 2. Rank order of vaccines for each individual test system

Test system	Rank order of vaccines [a]					
	A	B	C	D	E	Control
1	2.5	4	1	5	2.5	6
2	1	6	3	2	5	4
3	1	3	2	6	4	5
4	1	4	2	5	3	6
5	2	3	1	5	4	6
6	3	1	2	5	4	6
7	3	1	2	4	6	5
8	3	5	1	4	2	6
9	2	4	5	1	3	6
10	1	5	4	2.5	2.5	6
11	1	6	4	2	3	5
12	2	3	1	4	6	5
13	1	2	3	5	4	6
14	1	4	2.5	5	2.5	6
15	1	6	3.5	2	3.5	5
16	1	4	5	3	2	6
17	4	6	3	5	1	2
18	1	2	5	3	4	6
19	1	5	4	3	2	6
20	6	3	5	1	2	4
21	1	2	3	5	4	6

[a] When two vaccines were equally effective in a given test system, each was assigned the same rank.

RESULTS

The basic question asked in this experiment was: does the potency ascribed to an immunogenic material vary with the test system in which it is evaluated? Table 2 gives the ranking of the vaccines for each test system and Table 3 gives the analysis of variance for these data. The main interest in the statistical analysis of this experiment is in the interaction terms of the analysis of variance because these terms indicate whether laboratories or animal models, or both, are consistent in the ranking of the vaccines. The F ratios for the " animal model within laboratories by treatment " terms are unconfounded tests of whether the different animal models within a particular laboratory rank the vaccines in the same order. On the other hand, the " laboratory by treatment " interactions are confounded, i.e., are not a pure test of whether different laboratories agree on the rank order of the vaccines, since every laboratory did not use the same animal models. In the analysis of variance shown in Table 3, the first logical contrast (p) compares the non-vaccinated group with the average of the 5 vaccinated groups. The second term (q) compares the non-living vaccine given in adjuvant with non-living vaccine given in saline. The third term (r) contrasts the average of the live vaccines with the average of the dead vaccines. The fourth term (s) compares the two live vaccines BCG and *Myco. avium*, and the last term (t) compares the defatted killed vaccine with the average of the non-defatted killed vaccines.

All 10 interaction terms are significant at the 0.001 level. This indicates significant disagreement between the test systems on the rank order of the 5 vaccines. Because all the interaction terms were

Table 3. Analysis of variance of 21 assays of a series of vaccines

Source	Degrees of freedom	Mean square	F ratio [a]
Test system	20		
Laboratory	8		
Animal model: laboratory	12		
Treatment	5		
(*p*) 5 (NV) *versus* (A+B+C+D+E)	1	21.6017	510.68 *
(*q*) C *versus* D	1	1.2387	29.28 *
(*r*) 3 (A+B) *versus* 2 (C+D+E)	1	1.5135	37.78 *
(*s*) A *versus* B	1	8.9840	212.39 *
(*t*) 2 (E) *versus* (C+D)	1	0.0186	0.44
Test system x treatment	100		
Laboratory x treatment	40		
(*p*)	8	0.2535	5.99 *
(*q*)	8	0.7716	18.24 *
(*r*)	8	0.7415	17.53 *
(*s*)	8	0.8681	20.52 *
(*t*)	8	0.5665	13.39 *
Animal model: laboratory x treatment	60		
(*p*)	12	0.3730	8.82 *
(*q*)	12	0.7618	18.01 *
(*r*)	12	0.4272	10.10 *
(*s*)	12	0.4680	11.06 *
(*t*)	12	0.2282	5.39 *
Within	1572	0.0423	
Total	1697		

[a] * = $P<0.001$

significant, it is necessary to calculate the F ratios for the 5 comparisons for each test system (Table 4). Considering only the significant F ratios (F greater than 3.85), the underlined values of the F ratio indicate that the two components in the comparison were ranked in reverse order to those that are not underlined. For example, the comparison of the non-living vaccine in saline with the non-living vaccine in adjuvant (C against D) shows that all but 2 test systems detected a significant difference between C and D but 12 reported C to be more potent than D and 7 reported D to be more potent than C. Of the 5 logical contrasts for the main effects, 4 (*p*, *q*, *r*, *s*) are statistically significant at the 0.001 level, thus indicating that there is some general degree of agreement among the 21 test systems for these 4 logical contrasts.

This experiment demonstrates that the rank order of these vaccines was influenced at least by the animal-model component of the test system.

Table 4. F values for logical contrasts of vaccines for each individual test system [a]

Test system	NV *versus* V	C *versus* D	Live *versus* dead	BCG *versus* avian	2E *versus* C+D
1	45.54	43.08	1.62	4.60	0.19
2	1.16	2.29	0.13	75.90	17.84
3	11.69	43.08	28.95	13.41	0.19
4	45.54	19.49	6.14	25.50	0.64
5	45.54	43.08	2.24	2.29	4.17
6	45.54	19.49	18.24	13.41	0.64
7	11.69	8.42	28.95	13.41	34.85
8	45.54	25.51	17.66	8.42	0.19
9	45.54	43.08	0.08	8.42	0.19
10	45.54	4.60	0.27	43.08	1.53
11	11.69	8.42	1.30	75.90	0.00
12	11.69	25.51	6.79	2.29	51.04
13	45.54	8.42	38.73	4.62	0.00
14	45.54	13.38	6.14	25.50	4.46
15	11.69	4.87	1.30	75.90	1.62
16	45.54	8.42	6.14	25.50	11.71
17	11.69	8.42	28.95	13.41	34.85
18	45.54	8.42	38.73	4.64	0.00
19	45.54	1.93	0.27	43.08	6.49
20	1.16	43.08	25.86	25.50	1.71
21	45.54	8.42	38.73	4.62	0.00

[a] The underlined values are inversions. Critical $F_{0.05} = 3.85$.

DISCUSSION

The primary purpose of this work was to compare different methods used for assay of tuberculosis vaccines. The comparison was made by using a standard series of vaccines.

Requirements for a series of vaccines for such a comparison were stability and a potency ranging from no protective effect to the highest level of protection in experimental animals. BCG was chosen as the vaccine most likely to yield the highest level of protection. On the basis of reports by Palmer & Long (1966), *Myco. avium* was selected as a live vaccine that would give an intermediate level of protection. The extraction residue vaccine was selected on the basis of experiments that indicated that 50 μg of the vaccine in adjuvant gave to guinea-pigs a level of protection about half that provided by BCG. This evaluation was based on enumeration of the micro-organisms in the lung and spleen 5 weeks after infection and on survival time.[1] Preparations of formaldehyde-killed H37Rv with and without adjuvant were included because of the considerable disagreement in the literature (Smith, Grover & Wiegeshaus, 1968) about the immunogenicity of whole killed mycobacteria.

One objective of current research in experimental tuberculosis is to determine the relative immunizing potency of different preparations of the tubercle bacillus: living attenuated vaccines, killed vaccines, and various fractions. Because the mechanism of immunity in tuberculosis is unknown, there is no method to detect the immune state without com-

[1] Unpublished observations, D. W. Smith et al.

paring the response of vaccinated and control animals to infection. Comparison of the results obtained in different animal test systems requires that the potency of vaccines be independent of the test system in which it is determined. That this requirement has not been met can be seen from the results of this collaborative experiment, which demonstrates that the apparent potency of some vaccines is influenced at least by the animal model component of the test system.

The randomness of the ranking of potencies of the vaccines in the different test systems in this experiment indicates that the potency of a vaccine is a function of the animal model in which it is evaluated. Until the critical dependent and independent variables of the animal model are identified and their interactions determined, and until the importance of certain intra-laboratory variables is determined, it is not possible to compare the potency of vaccines evaluated in different test systems.

RÉSUMÉ

ÉVALUATION COLLECTIVE DES SYSTÈMES D'ÉPREUVE SERVANT À MESURER L'ACTIVITÉ DES VACCINS ANTITUBERCULEUX

Les problèmes que pose l'appréciation du pouvoir immunogène des vaccins antituberculeux ont retenu l'attention d'un groupe de conseillers de l'OMS qui, en 1964, a constaté le manque d'uniformité des résultats obtenus dans les divers laboratoires travaillant dans ce domaine. Une revue de la littérature consacrée à cette question depuis une dizaine d'années conduit à la même conclusion. On attribue généralement ces discordances aux procédés différents de préparation des vaccins, mais on peut aussi soupçonner qu'elles sont dues à la diversité des méthodes employées pour évaluer leur activité.

La présente étude a été conçue en vue d'établir si les variations dans les méthodes d'épreuve ont une influence sur les résultats des mesures du pouvoir immunogène des vaccins antituberculeux. Les neuf laboratoires de huit pays qui ont pris part aux recherches ont reçu une même série de préparations: *a*) deux vaccins vivants lyophilisés (une souche de BCG et une souche de *Mycobacterium avium*) et *b*) trois vaccins tués: deux vaccins à base de *Myco. tuberculosis* inactivés par le formol, en suspension saline ou en suspension dans l'huile minérale (Arlacel A), et un résidu d'extraction de *Myco. tuberculosis* en suspension dans l'Arlacel A. Les laboratoires ont été invités à injecter aux animaux d'expérience une dose déterminée de chaque vaccin et à procéder, huit semaines plus tard, à l'infection d'épreuve par une souche virulente de *Myco. tuberculosis*. Tous les autres paramètres propres aux divers modèles animaux devaient être identiques à ceux normalement utilisés par les laboratoires: espèce animale, voies utilisées pour la vaccination et l'infection d'épreuve, intervalle entre cette dernière et le sacrifice de l'animal, etc. Les laboratoires ayant l'expérience de plusieurs modèles animaux ont été encouragés à employer différentes méthodes. Au total, 21 systèmes d'épreuve ont été évalués.

Pour apprécier l'efficacité des vaccins, les laboratoires avaient eu recours à des indices variables mesurant la réponse de l'hôte animal à l'infection: temps de survie, indice splénique, numération des mycobactéries dans les tissus splénique et pulmonaire, etc. Dès lors,l a comparaison de leurs résultats n'a été possible qu'en adoptant comme critère un classement préalable des vaccins basé sur leur activité apparente. Toutes les données reçues ont été soumises à une analyse de variance selon la méthode des blocs complets casualisés. Elle a fait ressortir une très nette discordance des résultats fournis par les 21 systèmes d'épreuve étudiés en ce qui concerne l'étalonnage des vaccins considérés isolément ou groupés selon diverses combinaisons. D'autre part, des divergences se sont manifestées dans les données obtenues par un même laboratoire après application de différentes méthodes d'épreuve.

On conclut qu'aussi longtemps que n'auront pas été identifiées les variables propres à chaque modèle animal, et déterminées leur importance et leurs interactions, il n'est pas possible de comparer l'activité des vaccins antituberculeux mesurée à l'aide de systèmes d'épreuve différents.

REFERENCES

Fregnan, G. B. & Smith, D. W. (1963) *Amer. Rev. resp. Dis.*, **87**, 877-888

Harter, L. H. (1961) *Biometrika*, **48**, 151

Li, J. C. R. (1964) *Statistical inference*, Ann Arbor, Edwards Bros., Inc., vol. 1, p. 252

Palmer, C. E. & Long, M. W. (1966) *Amer. Rev. resp. Dis.*, **94**, 553-568

Smith, D. W., Grover, A. A. & Wiegehaus, E. H. (1968) *Fortschr. Tuberk.-Forsch.*, **16**, 191-227

Ungar, J., Muggleton, P. W., Dudley, J. A. R. & Griffiths, M. I. (1962) *Brit. med. J.*, **2**, 1086-1089

Annex

LIST OF COLLABORATING INSTITUTIONS

CZECHOSLOVAKIA

Dr L. Šula, Dr J. Galliová, Dr J. Průchová & Mrs J. Šulová
Tuberculosis Research Institute
Prague

DENMARK

Miss K. Bunch-Christensen & Miss A. Ladefoged
BCG Department
Statens Seruminstitut
Copenhagen

FRANCE

Dr F. Lévy & Dr G. Conge
International Children's Centre
Paris

HUNGARY

Dr I. Földes
Koranyi National Institute of Tuberculosis
Budapest

ROMANIA

Dr C. Oprescu
Institute of Microbiology, Parasitology, and Epidemiology
Bucharest

UNION OF SOCIALIST SOVIET REPUBLICS

Dr T. Jablokova
L. A. Tarasevich State Control Institute of Medical Biological Preparations
Moscow

UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND

Dr P. W. Muggleton
Glaxo Laboratories, Ltd
Greenford
Middlesex

UNITED STATES OF AMERICA

Dr E. Ribi & Dr R. Anacker
National Institutes of Health
Rocky Mt. Laboratories
Hamilton, Mont.

Professor D. W. Smith, Mr E. Wiegeshaus, Mr G. Harding, Mr D. McMurray & Dr A. A. Grover
University of Wisconsin
Madison, Wis.