

原 著

ツベルクリン反応検査に伴ういくつかの誤差要因の分析

森 亨・高井 鎌 二

結核予防会結核研究所

受付 昭和 48 年 6 月 20 日

ANALYSIS OF ERRORS AND VARIATIONS IN TUBERCULIN TESTING*

Toru MORI and Ryoji TAKAI

(Received for publication June 20, 1973)

A factorial experiment was performed in order to evaluate quantitatively the importance of technical errors in giving and reading tuberculin test. The analysis of variance showed that all of the involved sources of variation were statistically significant. These were; 1) among-subject (in this experiment, tuberculosis patient) difference, 2) between-injection difference in a subject (namely the variation in giving injection), 3) among-reader difference, and interactions. There was wide variation also in the judgement of whether a reaction had a so-called double erythema. Discussion was made on each of the errors. Considering that the tuberculin testing cannot be free from these errors, it was keenly felt necessary to standardize technique in the performance of the test through training of the personnels, and to take greatest care in interpreting the test-results.

I. 目 的

ツベルクリン（以下「ツ」と略）による皮膚反応検査は、個人・集団用に広く用いられ、わが国の結核対策上に占める位置は、一応確立したものとされている。

しかし、よく知られているように、この検査は多くの因子によつてその結果が影響される。そして、このことの不十分な認識が、時に臨床上あるいは疫学・評価のうえでの誤った判断の原因になつていることも指摘されている（たとえば¹⁾¹⁹⁾）。これらの因子（変動要因）のいくつかの基本的なものについて、その重要性を数量的に明らかにするのが、本研究の目的である。

II. 材料と方法

本研究の被検者の当研究所附属療養所に入所中の 25 人の結核患者である。各被検者の両側前腕内側の下 1/3

の部位の皮内に、両側とも PPD-s（強反応者用、ロット番号 P 1001、このロットは 0.05 mcg 用であるが、これを 5 倍に希釈して使用した）0.01 mcg/0.1 ml をいれ、leak-free の注射筒（Microstat®, Omega）で注意深く注射した。注射は全部 1 人の経験ある術者が行つた。

計測は注射後 48 時間目に、簡単のため発赤・硬結とも横径のみをキャリパーで測り、また二重発赤の有無は別に記載した（二重発赤ありとされた者の反応については、外発赤径を発赤径と定義した）。計測にあつたのはいずれも 5~20 年以上にわたり毎年 3,000 人以上についてツ反応計測を日常行つている医師 3 人、保健婦 1 人である。計測に際して、計者には用いたツ液の量、左右の腕に用いた抗原の違いの有無（当所では左右の腕に異なる種類のツ反検査を行うことはしばしばある）、また研究の目的等を、意図的に一切知らせず、ただでざる限

* From the Research Institute of Tuberculosis, Japan Anti-Tuberculosis Association, Matsuyama, Kiyose-shi, Tokyo 180-04 Japan.

り慎重に計測することのみを要請した。

各被検者は左右の腕の反応につき、3 回ずつすべての計測者から計測を受ける。この際同じ反応に関する前の計測結果が覚えられたいないように、計測の順序、被検者の流れに工夫をした。また発赤と硬結は完全に独立・別個に測つて、1 つの反応に関するそれらの計測値が影響しあわないようにした。計測作業は 2 人の係員が監督し、作業の流れの混乱や記載の誤りのないようにつとめた。

このようにして、1 人の被検者は；4 人の計測者×3 回の反復計測×2 カ所の反応×2 種の（発赤・硬結）反応=48、と合計 48 コの計測値を与えられることになる。

III. 分析結果

① 全体のツ反応分布：被検者にはそれぞれのツ・アレルギーの計測値として、発赤・硬結につき 24 コの値が与えられるが、これらの平均値は、被検者のツ反応の真の値に最も近いものと考えられる。この値の全員についての分布は図 1 のようになる。ともにほぼ定型的な正規分布であり、以下の分布の正規性を前提とする分析に支障ないことが確かめられた。

② 二重発赤判定の一致・再現性：図 2 は、被検者の

Fig. 1. Distribution of Mean Size of Reaction (Based on all the readings of both tests in a subject)

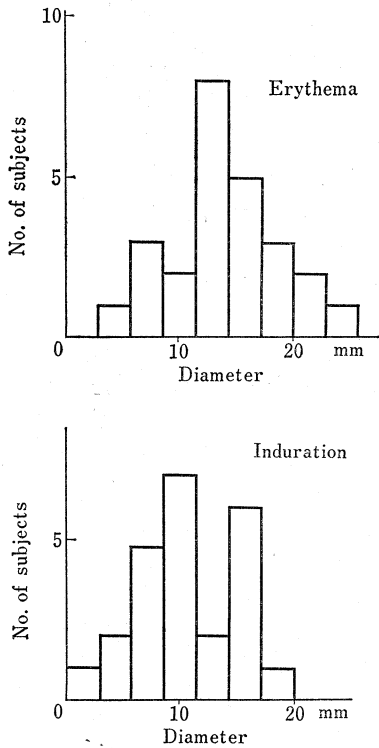


Fig. 2. Reproducibility of Interpretation of a Reaction with/without Double Erythema

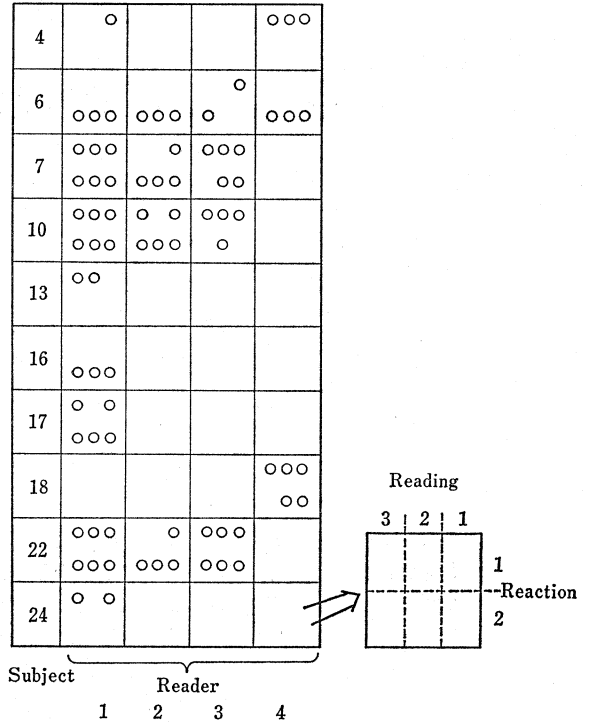
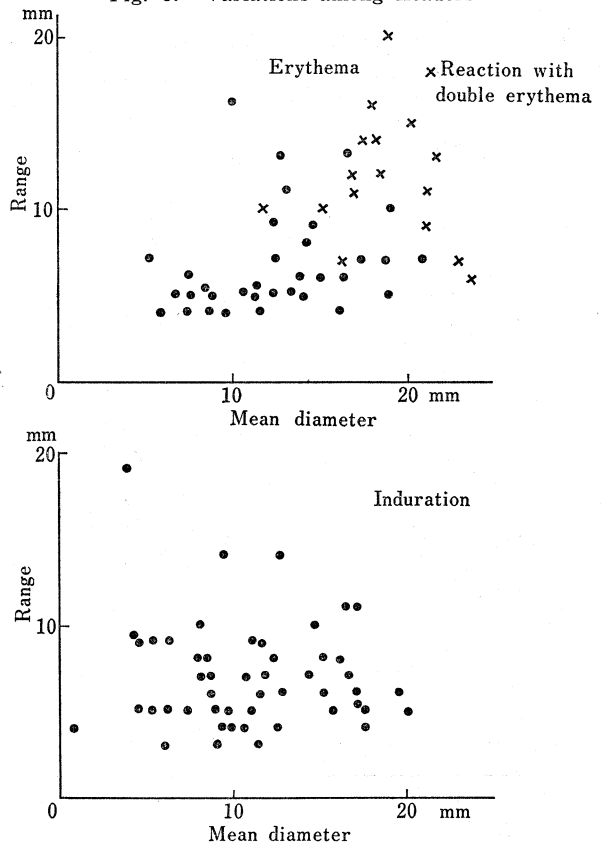


Fig. 3. Variations among Readers

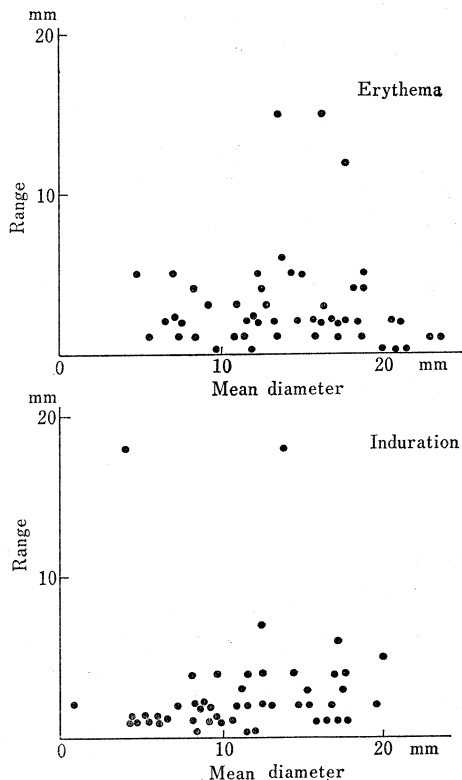


延べ 24 回の発赤計測のうち 1 回でも二重発赤ありとされた者 10 人について、左右の反応別、計測者別、計測反復別に、一致・再現性をみたものである (図中○印で、二重発赤ありと判定したことを示す)。10 人中、全計測者が各回とも一致して二重発赤ありとした反応は皆無、同一計測者でも計測のたびごとに一致して「あり」としている例は少なく、同一被検者においても左右の反応ごとに二重発赤の有無が一致しないことが多い、等の事実が知られる。計測者間の比較をみると、計測者 2, 3 は比較的同傾向の判定をしているのに対して 1 は読みすぎ、4 は過少評価あるいは、他の 3 人とは少し異なる基準で判定していることがうかがわれる。

③ 全計測間のバラツキ：各反応の 12 コ内のバラツキをより直観的にみるために、これらの計測値中の最大・最小値の差 (範囲) で表現し、これを 12 コの値の平均値 (この反応の大きさの真の値に最も近いと考えられる) に対してプロットしてみた (図 3)。二重発赤ありの判定を一度でもされた者については、その判定の不一致性のため当然範囲は大きくなると考えられるので、図では特に×で示してある。同一の反応を計測しても、計測する人が変わったり、反復して測つたりする際には常に同じ値が得られることはなく、発赤で最大 20 mm、平均 8.3 mm、硬結でそれぞれ 19 mm、6.9 mm というかなり大きな幅のバラツキが生じることがわかる。平均値 (上に定義した) と範囲の相関係数は、発赤、硬結についてそれぞれ 0.458**, 0.094 (n. s.) であった。

④ 計測者内のバラツキ：ある 1 人の計測者の各反応あたり 3 コの計測値につき③と同様のことをみた (図 4)。横軸は便宜的に③と共通のものを用いた。バラツキの幅は③のよりも当然小さいが、それでも発赤で範囲の最大値は 15 mm、平均値は 3.0 mm、硬結で 18

Fig. 4. Variations Within a Reader



mm、2.6 mm であった。反応の平均径と範囲の相関係数は、発赤、硬結についてそれぞれ、0.01 (n. s.), 0.05 (n. s.) であった。

⑤ 分散分析による各変動要因の検討：この実験は 3 回の繰返し計測をもつた 25 人の被検者因子と 4 人の計測者因子とをかけ合せた二元配置に、被検者一注射のかけ合せの一元配置 (50 コの注射が被検者によつて層化さ

Table. Results of Analysis of Variance

Sources of variation	df	Erythema					Induration				
		SS	MS	F	$\hat{\sigma}$ (mm)	VC(%)	SS	MS	F	$\hat{\sigma}$ (mm)	VC(%)
Subjects S	24	13,224	551.0	13.4**	4.63	32.8	11,939	497.5	19.5**	4.46	40.3
Readers R	3	1,251	417.0	20.5**	1.63	11.6	1,092	364.2	48.4**	1.54	13.9
Injections I(S) §	25	538	21.5	2.1*	0.98	7.0	454	18.2	4.8**	1.09	9.9
Interaction SR	72	1,464	20.3	2.0*	1.31	9.3	546	7.6	2.0*	0.79	7.1
Interaction RI(S)	75	753	10.0	2.1*	1.32	9.4	287	3.8	1.4*	0.63	5.6
Residuals +	400	1,938	4.8		2.20	15.6	1,055	2.6		1.62	14.6
Total	599	19,167					15,375				

§ : See text.
 + : Errors due to repeated readings of a reaction by a reader
 df : Degree of freedom
 SS : Sum of squares
 MS : Mean square
 $\hat{\sigma}$: Square root of "componet of variance", see text.
 VC : Coefficient of variation, see text.

れた形で)が組み合わされた形の要因配置実験とみることが出来る(各因子の効果はともに変量型とした。)¹⁾⁹⁾。ここで“注射”因子には、50コノ注射中の個々の技術の変動と、被検者体内注射部位差による変動の双方が関係しているが、ふつうは前者のほうが大きいと考えられるので、単に注射による変動と名付けた。さらにこの注射効果は、被検者内での2つの注射間の変動を全被検者にわたって合成したものであつて、50コノ注射全体にわたる変動そのものではない。

以上のように考えてまとめた分散分析表が次の表である。表に掲げたように、考えられる変動要因としては、(i)被検者間、(ii)計測者間、(iii)注射(被検者内)間、(iv)(i)と(ii)の交互作用、(v)(ii)と(iii)との交互作用、および(vi)残差(反復効果)がある。F検定によれば、(i)~(vi)の変動要因の効果はすべて有意である。

各変動要因の平均平方の期待値の式から変動要因による効果の母分散(の平方根)の推定値($\hat{\sigma}$) (分散成分の平方根)を算出したものが表の値である。上の分散分析でのF検定でこれらの値(残差に対するものを除く)が0でないことが有意とされているのであるが、これらが一応計測値の変動に対する各要因の効果の大きさの尺度となりうるものと考えられる。

被検者因子に対する $\hat{\sigma}$ の値が最大なのは当然である。発赤・硬結では計測値そのものの大きさが全体的に異なるから、それらの全計測値の平均値(それぞれ14.1mm, 11.1mm)との比(変異係数, 表中VC)をみると、交互作用、残差に対するもの以外は、硬結のほうにより大きくなっている。

IV. 討 論

ツ反応の計測値に影響する因子としては、生物学的因子のほかには、技術的なものとして主として、使用する抗原の特性、注射に関係したものと、計測に関するものが考えられる。

すなわち、ツベルクリンの種類、投与量、溶液の容積(あるいは濃度);注射の深さ、部位;計測者内、計測者間のバラツキ、ないし傾向(バイアス);等等である。これらの1つ1つに関してはGuld⁴⁾¹⁵⁾、Wijsmuller⁵⁾らが詳しい検討を行つている。日本では主としてOTを用いた検査に関して、柳沢ら²⁰⁾によつてまとめられている。この研究はそれらがPPDについて実際どの程度の重大さをもっているかを見るものであつた。

ツ反応検査がいかに幅の大きいバラツキに支配されているかを知るには図3, 4をみれば十分であろう。ここでは直観に訴える目的で範囲をバラツキの尺度として採用しているが、興味深いのは図3の発赤の場合を除いて、バラツキは各反応の大きさ(平均値)と相関しない

ことである。図3の発赤の場合については、二重発赤ありと判定されたもの(全員・全回一致でない場合が多く、このため範囲が大きくなり、同時に平均値も大きくなる。)が、グラフの右上に集中することによると考えられる。したがつて、上のバラツキと反応の大きさの無相関ということは(二重発赤を除いていえば)、一般的にいえるのではないか、と思われる。Guldら⁴⁾およびHorwitzら⁶⁾も一般に中等度以上の反応では反復検査、反復計測の値のバラツキは小さく、弱い反応のときには大きくなるとしているが、0.01mcgを使用しているとはいえ、この研究では結核患者でみているので、前者の場合に該当しているものといえよう。逆にいえば、BCG接種後にみられる弱い反応を計測する場合には、バラツキはこれよりも大きくなると考えるべきである。

質の判定としての二重発赤有無の判定も決してclear-cutに行いうるものでないことは図2によつて明らかである。これは0.01mcgという使用したツベルクリン量の小ささのため、典型的な二重発赤ができにくかつたためであろうが、一方主観的要素の強い判定の際には避けられない不一致でもあろう。

分散分析の結果によつて示された諸種の効果の母分散推定値のうち最大のもの、当然被検者間のバラツキに対するものであつた。この値は、おおざっぱにいえば、被検者の生物学的背景因子がこれと同じくらいの集団で、注射・計測の誤差のない検査が行われたと仮定した場合、硬結の平均が11mmとすれば、被検者全体の95.5%が約2mmから20mmの間に入るであろう、ということの意味する。これだけの幅の変動を構成する1つ1つの生物学的因子については今回は触れない。

次に大きな幅の効果をもつ変動要因は、表中で計測誤差と呼んでいるものである。これは同一計測者が同一反応を反復計測したときのバラツキを反映している。硬結についてNissen-Meyerら⁷⁾はこれを1.3mmとし、Bearman⁸⁾の成績から著者が算出したものでは1.5mmとなり、いずれもわれわれのものから遠い値ではない。変異係数が硬結よりも発赤の場合に大きいのは、二重発赤例の不一致が響いているからであろう。われわれは「二重発赤あり」の反応についてはいわゆる外発赤径を発赤径としたが、もし内発赤径を発赤径にすれば、この誤差はより小さくなるものと思われる。「交互作用」についても同様である。

“計測者”因子は、各計測者の、1つの定型的な反応を計る際の傾向(バイアス)の程度を示していると解釈しうる。Bearmanの成績からは、硬結に対して1.7mmと計算でき、これもわれわれのものに近い。一方、定型的でない反応を含む個々の反応を計測する際の計測者の傾向の違いによる変動は“注射×計測者”、“計測者×被検者”という交互作用として表現されるものと解釈でき

る。これらの θ が発赤についてかなり大きいのは、二重発赤の有無の判定に計測者の個人差があつたこと、しかも注射の仕方(ないし部位)によつて同一被検者でも二重発赤が生じたり生じなかつたりすることがあること、等によると考えられる。硬結についても、非定型的な硬結の場合に、これを無視してしまう者と、注意深く硬結を触知する者等の差が、これらの交互作用の変動効果を統計的に有意のものにしているであろう。「方法」に述べたように、本実験では硬結・発赤を全く別個に計測したが、その結果、同一反応についての発赤径よりも硬結径の計測値が大きくなるような組合せになることもまれでなかつた。しかし、これは純粹の計測という点で、先入観のない、バイアスのない計測としては当然のことである。

“注射”因子による変動は、「分析」で述べたように、被検者内部位差をも含んでおり、これらを分離したものとして検定するには、別のより精密な実験を設計しなければならない。この実験では、注射される液量を正確に0.1 ml とするために、ピストンと筒の間(すり合わせ部)、筒と針の接合部からの液の加圧による漏出⁹⁾がはるかに少なくなるよう工夫され、また目盛も細かく刻まれた注射器を用いた。この種の注射器は日本でもつと普及してもよいのではないだろうか。

また注射の作業自体も終始1名の技術者がかなり慎重に行つたものである。日常業務でのこの因子による変動はもつと大きいものと思われる。このことは、何度も触れるが、他のすべての変動因にも共通するものであること、さらに対象集団が変れば、誤差そのものの大きさは当然変るものであることを、重ねて強調しておきたい。

では、以上のような誤差やバイアスはどのようにしたら克服できるか。Bearman らは、かなり規格化した技術をもつて行つたにもかかわらずマンツウ法にはある程度の誤差は避けられないからといつて、Heaf 法、tine 法等の新しい方法がそれよりよいとすることは楽観的すぎると警告している⁸⁾。著者らは、現状としては、注射・計測の訓練による規格化、および実施の慎重さにまつのが最善と考えている。訓練の方法としては、標準計測者との読み合せ、被訓練者自身の繰返し計測による一致性の評価を繰返していくやり方が諸外国ではよく行われている¹⁰⁾¹¹⁾。事実このようにして規格化された技術は驚くほどの再現性を示す。たとえば、デンマークのいくつかの学童の群に BCG 接種直後とその5年後にツ反計測をしたところ、学童1人1人の2回の計測値は非常にまちまちであつたが、群全体の平均値は各群について1 mm 以下の違いしかみられなかつたという⁶⁾。慎重さについてはいうまでもないが、たとえば、チェコの tuberculosis nurse (ツ反注射・計測、BCG 接種、喀痰採取等結核関

係の専門看護婦)は、1時間に計測できる最大の数は、100名を越えないといっている¹²⁾。

同時に重要なのは、これらの計測値に、一定の基準(限界)をあてはめて、個人に関するなんらかの判定を下す場合の誤差に対する考慮である。このことは、たとえば、BCG 接種後の再接種対象の選定¹⁾、個人に関する陽転・(いわゆる)陰転の診断¹³⁾¹⁴⁾などについて問題になつているし、さらにサルコイドーシス、非定型抗酸菌症の診断や研究についてもいえる。

種々の因子によるバラツキの大きい検査の結果を集団にでなく個人に適用することの危険性については、ツ反応以外にもすでにいろいろ論じられてきたところである(たとえば、文献^{16)~18)})。

最後につけ加えたいのは、旧来の OT からより特異性の高い精製ツベルクリン PPD を用いることで抗原の側に厳密さを期しているわが国としては、その実地使用面に関しても、精確さを確保する体系(技術評価・訓練制度等)の確立が要求されてもよいのではないだろうか、ということである。

(本稿の論旨は第47回日本結核病学会にて発表した。)

本研究に計測者としてご協力をいただいた結核研究所附属療養所医師塩沢活、中村健一両先生、保健婦諸星正枝女史、およびご助言、ご校閲をいただいた柳井晴夫(東大疫学)、豊川裕之(同上)、島尾忠男(結核研究所)の諸先生に感謝します。

文 献

- 1) 東義國：結核，45：213，1970。
- 2) 鳥居敏男 他：医学生物学のための推計学，東大出版会，1969。
- 3) Snedecor (畑村又好ほか訳)：統計的方法，岩波，1966。
- 4) Guld, J.: Acta Tuberc. Scand., 28：222, 1953.
- 5) Wijsmuller, G.: Selected Papers (The Royal Netherlands Tbc Assn.), Vol. 6, 1963.
- 6) Horwitz, O. et al.: Bull. W.H.O., 47：49, 1972.
- 7) Nissen-Meyer et al.: Public Health Report. 66：561, 1951.
- 8) Bearman, J. et al.: Am. Rev. Resp. Dis., 90：913, 1964.
- 9) Guld et al.: Acta Tuberc. Scand., 30：16, 1955.
- 10) Horwitz, O.: WHO/TB/Techn. Information/66, 48.
- 11) Teruel, J.R. et al.: Tubercle, Lond., 50：350, 1969.
- 12) Štaštna 個人的連絡。
- 13) 伊波茂雄 他：結核，47：345, 1972.
- 14) Wilson, L.M.: Am. Rev. Resp. Dis., 102：

- 282, 1970.
- 15) Guld, J.: Acta Tuberc. Scand., 30 : 16, 1955.
 - 16) Pickering, G. (勝沼晴雄ら訳) : 本態性高血圧症, 医師薬出版, 1964.
 - 17) Oldham, P. D.: J. Chron. Dis., 15 : 969, 1963.
 - 18) 福田安平 : 脳卒中・心発作の疫学的研究, 労働医学研究会, 1970.
 - 19) WHO Tuberc. Research Office ; Bull. W.H.O., 12 : 123, 1955.
 - 20) 柳沢謙他 : ツベルクリン反応, 金原出版, 1955.