1° TEMA

M. J. DAY

Royal Victoria Infirmary
Newcastle upon Tyne, England

Statistical Methods in the Analysis of Radiotherapy Results

Introduction

Possibly by way of introduction some comment is called for on the fact that our statistical theme should be introduced by a physicist such as myself. Of course I cannot claim to read the mind of the Organising Committee, but I am sure that they were well aware that several eminent medical physicists have indeed made important scientific contributions in this field. One such physicist in a lighter moment admitted that statistics was often thought of as a branch of cosmetics — something applied superficially at the last moment in order to improve the presentability of an otherwise almost perfect result — though he went on to deplore this practice.

In deciding to include this topic on the Agenda — for which personally I am glad — the Committee may perhaps have had two other considerations in mind. Firstly, I think many physicists have noticed a tendency among their clinical colleagues to shy away from the more mathematical and technical aspects of applied statistics. At this juncture some instinctive defence-mechanism seems to suggest that a brief consultation with the physicist might be helpful. Of course he isn't really the right man to go to for this kind of help, but at least he is on the spot, he is interested in the same problems and he may be expected to feel more at home with statistical ideas and the mathematical calculations which are involved.

At the same time medical physicists, from their own inside view of the treatment process, often see possible applications of statistical methods in the clinical field. I think this is a natural consequence of the active participation of the physicist in the treatment of the patient. The physicist's enthusiasm for quantitative methods is thereby confronted with the challenge of biological variability. The positive response to this challenge involves statistics - indeed the development of modern statistical ideas by Pearson, Fisher and many others was largely stimulated by the needs of biological research. Because my own work in medical physics has been chiefly in the traditional field of radiotherapy I shall illustrate my talk almost entirely from radiotherapy statistics.

The Probabilistic Approach

As we have seen in earlier sessions, physics made two important contributions to the development of radiotherapy as an organised system of treatment. First came the precise definition of the concept of dose and second the experimental technology of dosimetry. The establishment of the fundamental concept made quantitative dosimetry possible, with consequent advances both in techniques and in ideas. By analogy, what is the fundamental concept when we are concerned with the effect of a proposed treatment on a particular patient? In the context of this morning's discussion I want to suggest that it is the *probability of successful therapy*. Of course after the treatment has been carried out, probability may be replaced by knowledge. Unfortunately this is only of academic interest in so far as decisions affecting the treatment are concerned. At the relevant stage our knowledge can only be in terms of probabilities, and I believe it is exactly in these terms that the mind of the clinician works. Knowledge of the relevant probabilities is the key to the choice of the « best » treatment for an individual patient and the quantitative estimation of probability is the major aim of statistics.

The Survivorship Function

There are of course several criteria by which the success of a radiation treatment can be judged. For example, remission of symptoms, relief of pain, clinical or radiological resolution of the tumour, survival (with or without complete freedom from disease). Since the time factor is also involved we are concerned with a probability which I shall call P(t), which

is the probability of success at time t after treatment. In order to be specific I shall discuss P(t) as the survival probability, though I do not wish to imply that survival is the sole object of treatment.

The curve of P(t) against t has one dismal but mathematically advantageous feature — it is monotonically decreasing. (Fig. 1, showing P(t), the Survivorship Function). The quantity 1 — P(t), i. e. the probability that the patient will die before time t, is the cumulative distribution function of the individual survival times. Figure 1 illustrates the fact that the observed survival-rate is an estimate of P(t) which however cannot be known exactly.

If we differentiate P(t) we obtain the probability distribution function. Figure 2 illustrates the relationship between this theoretical curve and the histogram of actual survival times for the same group of cases as fig. 1.

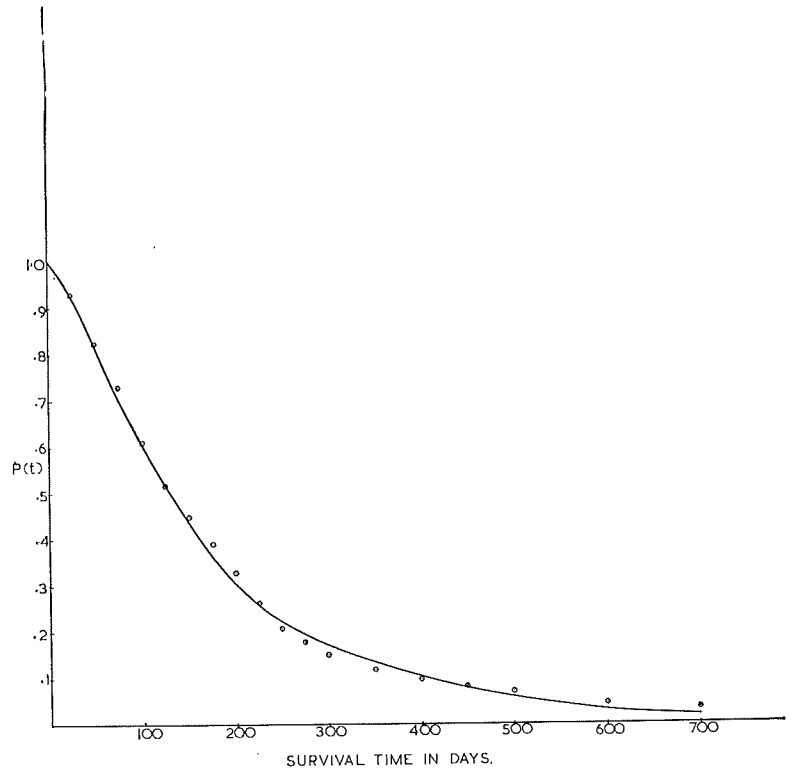


Fig. 1 - *The survivorship function. P(t) = probability of survival to time t days after treatment. Points are observed survival-rates for* 418 *cases of bronchogenic carcinoma.*

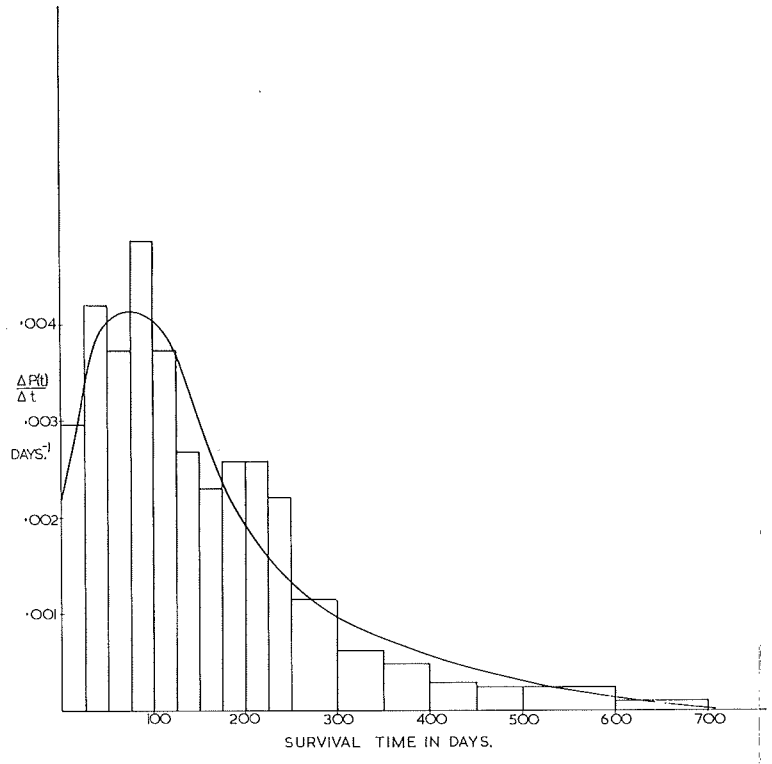


Fig. 2 - *The probability distribution function and its relationship to the histogram of observed survival times. Data form* 418 *cases of inoperable bronchogenic carcinoma.*

When we consider how P(t) can be measured we immediately encounter a difficulty because in principle it can only be done by observing the outcome of the same treatment on a group of identical patients. Since patients differ we must assume that their survival probabilities differ. Hence in effect we can really only estimate $\bar{P}(t)$, the average survival probability of the group. Whatever the individual variations may be, it does seem reasonable to regard this average survival probability as a suitable index of the merit of a particular treatment. If this is accepted then the arguments outlined in the Appendix show that we can, for calculation purposes, regard the group of patients as homogeneous even though we know that this is not strictly true. The appendix gives the variances of two possible estimates of $\bar{P}(t)$. One of these estimates takes into account the variation of $\bar{P}(t)$ from one patient to another; the

other ignores such variation. Suppose for example that the overall group really consists of three equal sub-groups with survival probabilities 70%, 50% and 30%. Then the estimate obtained by pooling all the patients has a variance only about 10% larger than that obtained by estimating the three P(t) values separately and subsequently forming the average of these estimates. It is of course necessary that the variations between patients should not be so large as completely to invalidate the general treatment policy - e.g. it would not usually be reasonable to group together patients with and without metastases, since a uniform method of treatment would be inappropriate. At the same time it *is* reasonable to group together patients covering a wide range of ages. My point is es-

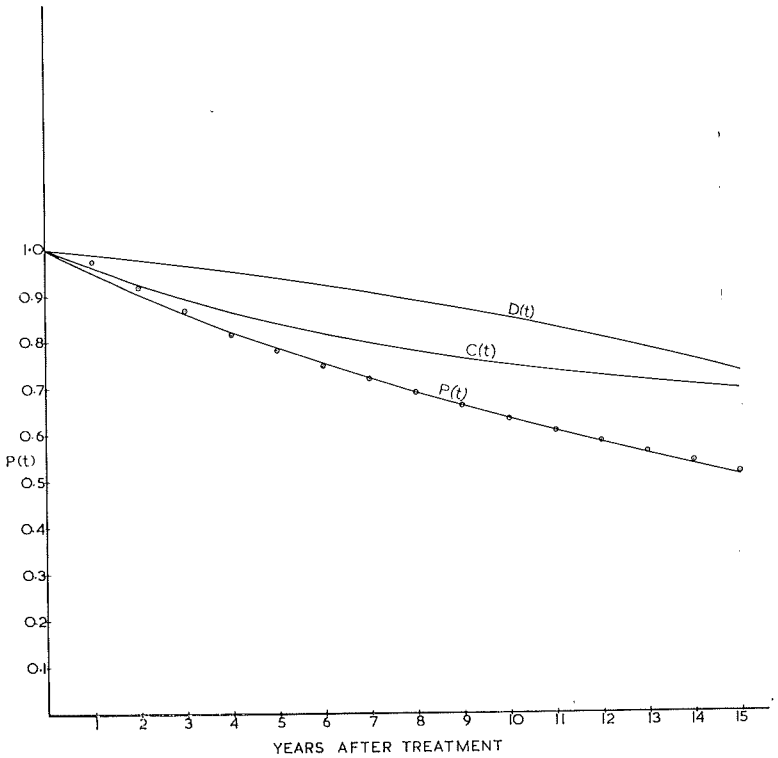


Fig. 3 - *Analysis of the survival function. Data for cancer of the breast without metastases from Berkson et. al. Jour. Amer. Stat. Ass.* 47-501 (1952).

$P(t) = C(t).\ D(t).$

Where: $P(t)$ = *Probability of Survival to time t.*

$C(t)$ = *Probability of survival to time t without dying of cancer.*

$D(t)$ = *Probability of survival to time t without dying of other disease.*

sentially that if we consider the individual patient in isolation P(t) becomes indeterminate; if we consider a small fairly homogeneous group $\bar{P}(t)$ can be estimated, but only roughly and if, at the other extreme we take a large and heterogeneous group, our estimate of $\bar{P}(t)$ becomes precise but rather useless. It is not very useful to know that the survival-rate for a well-defined category is say $40 \pm 30\%$, nor that the 5-year survival-rate for all patients with cancer is say 30%! Some compromise is therefore necessary, with the object of avoiding both the very large heterogeneous groups and also the unduly small groups, even if homogeneous.

Analysis of the Survivorship Function into its Components

The survival probability will mean more to us if we can separate the effects of cancer and of other possible causes of death. If these two effects operate independently then, as

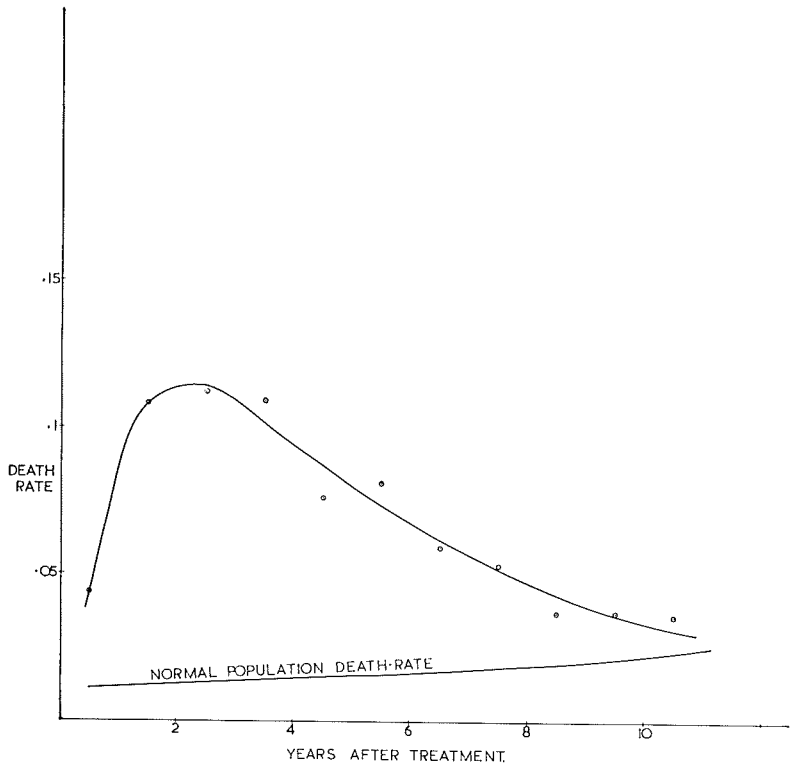


Fig. 4 - *Death rates in a treated series of operable breast cancer. Mc Whirter. R. Clinical Radiology XI,* 144 (1960).

shown in Fig. 3, the overall survival probability is equal to the product of the two separate probabilities. We use this formula in practice to deduce C(t) the cancer survival probability, in which we are chiefly interested.

Incidentally the death-rates corresponding to these two categories of mortality have very different mathematical forms. The cancer death-rate decreases with time after treatment (Fig. 4), whereas Gompertz pointed out in 1825 that the general death-rate increases exponentially with age. According to Haybittle, (Fig. 5) the cancer death-rate falls exponentially. As a

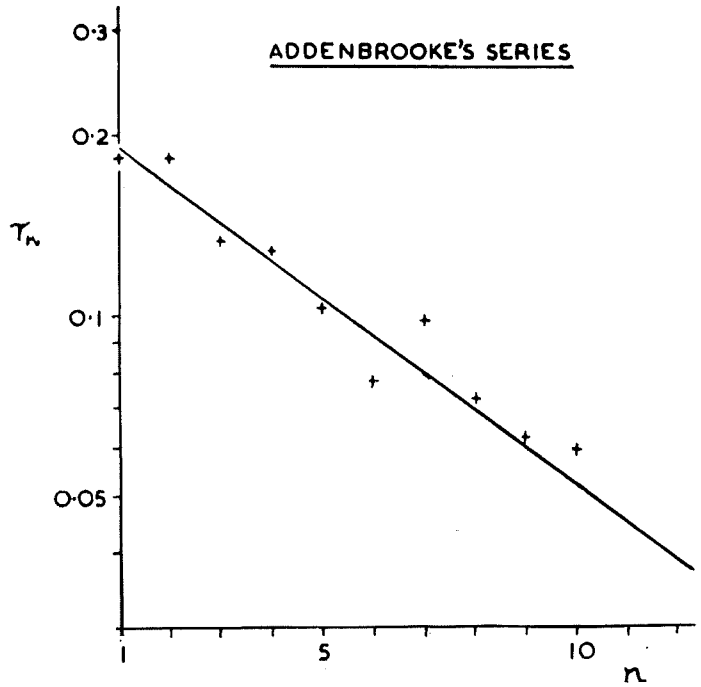


Fig. 5 - *Death-Rates in Cancer of the Breast.*
Reference: Haybittle (1959).

consequence the « normal » death-rate becomes important only when survival times are relatively long, and it is only in such cases that our assumption of independently-acting causes of death is liable to be seriously in error.

Table I, which gives data for cancer of the stomach, illustrates the importance of remembering the composite nature of the survival probability. Note that the crude survival rates

TABLE 1

CARCINOMA OF THE STOMACH: 5-YEAR SURVIVALS ACCORDING TO AGE

Age, years	*Patients*		*Lived 5 or more years after leaving hospital*		*Survival rate adjusted for normal death rate*
	Total	*Traced*	*Number*	*Survival rate, per cent*	
Less than 40	206	203	61	30.0	30.9
40-49	672	666	203	30.5	32.2
50-59	1,118	1.110	342	30.8	34.3
60-69	976	971	300	30.9	39.1
70+	240	239	70	29.3	50.2
Total	3.212	3,189	976	30.6	34.4

REFERENCE: Berkson et al. (1958).

[i.e. P(t)] show little variation with age. If however we consider the corrected survival-rates [i.e. C(t)] then there is a definite upward trend with age. These results thus conform to the general finding that tumours in older persons tend to be less malignant than those in younger people.

It is important that the estimation of P(t) shall not have to await complete results for every patient in the series. There are always a few patients « lost to follow-up » and furthermore it is very desirable that use should be made of all the available information, including that from patients who have been treated relatively recently. In such cases the patient is said to be « withdrawn from observation », and withdrawal is mathematically equivalent to a third cause of death. Valid estimates of P(t) based on such incomplete data are still possible, though inevitably less accurate than those based on the final data. However it is a real advantage that results are available perhaps several years earlier than would otherwise be possible.

Parametric Methods

In my view it is not generally realised that there are a number of alternative methods for estimating the survivorship function P(t). Let us take the *parametric methods* first. Under

this general heading we include all methods in which the survival curve is assumed to possess a definite mathematical form. For example the curve might be exponential, in which case it could be completely specified by the half-life. Alternatively we might wish to assume that the survival times were normally distributed, in which case two parameters (the mean and variance) would be necessary to specify the complete curve. In either case the survival curve which most nearly

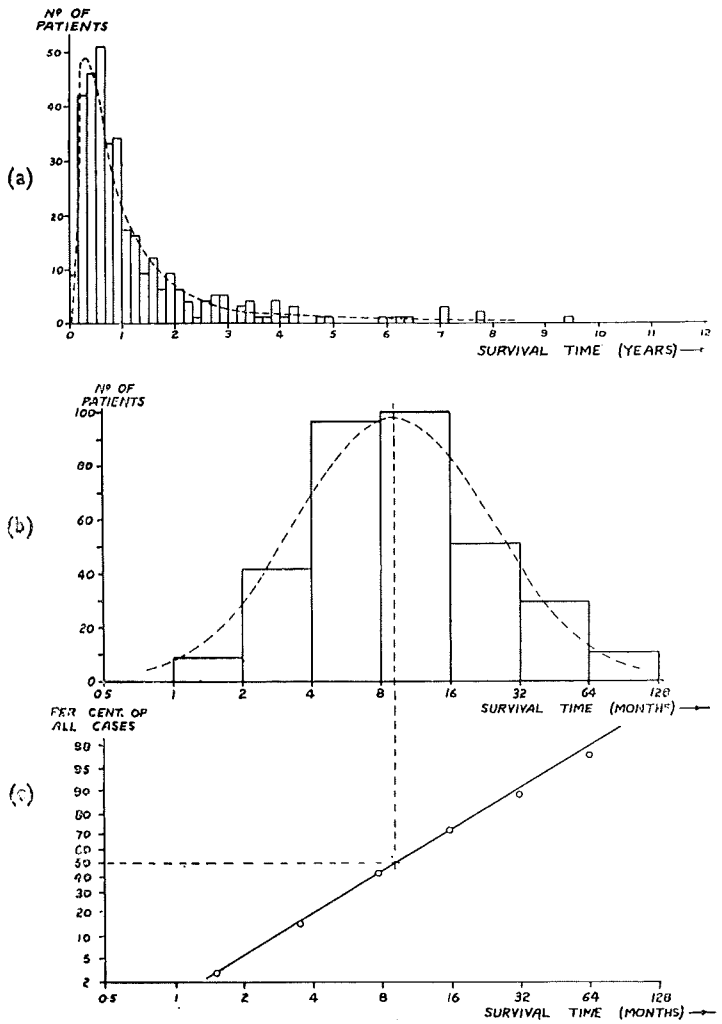


Fig. 6 - *Log-Normal Distribution of Survival Times.*
Reference: Boag et al (1950).

accorded with our data could be constructed by the Method of Maximum Likelihood. This is a general theory of great power, first described by Fisher in 1922 and applied to our problem by Boag in 1949. Boag's theory is based on his discovery that survival times in patients who died of cancer were very nearly longnormally distributed. As shown in Fig. 6., the ordinary histogram of survival times is very skew, but if we transform to a logarithmic time-scale the histogram becomes almost symmetrical. The normality of the distribution can be demonstrated by plotting the survival rates on a transformed « probability » scale as shown at the foot of this diagram. Using this method of graphing the data, the longnormal distribution gives a straight line from which the parameters can be estimated approximately. Fig. 7., which shows data of Sorensen

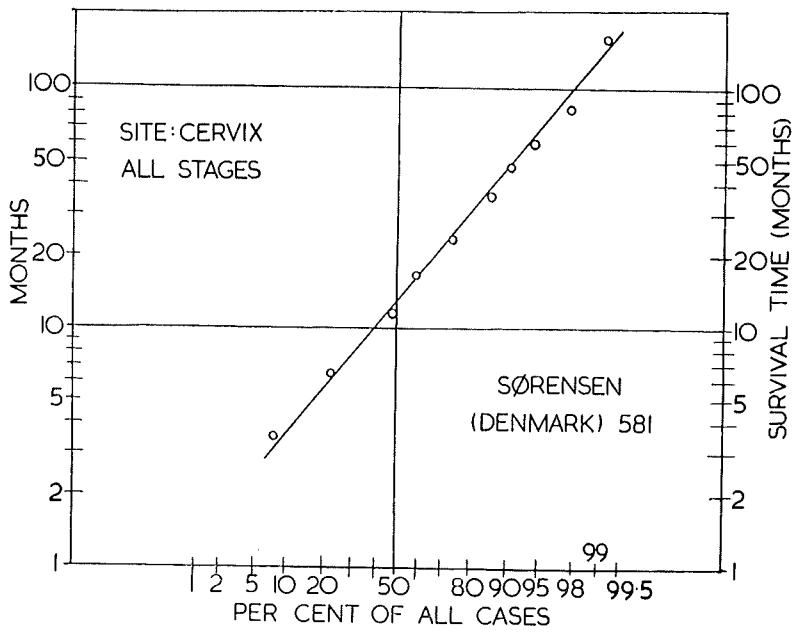


Fig. 7 - *Long-Normal Distribution in Carcinoma of the Cervix.*
Reference: Sorensen (1958), Boag (1960).

for carcinoma of the cervix uteri, is another good example of the doubly transformed survival curve. Boag's model assumes further that a proportion C of the patients are completely cured of cancer and are subject only to other causes of death.

We thus have three parameters to estimate and Boag developed a method for doing this. In practice certain computational difficulties arise unless the data are very extensive and for this reason it is usual to fix the variance so that there are only two unknown parameters.

A rather similar approach was proposed by Berkson, who suggested that the cancer survival curve was exponential, but that again a certain proportion of cases was subject only to the normal force of mortality. Figure 8 shows that this theory

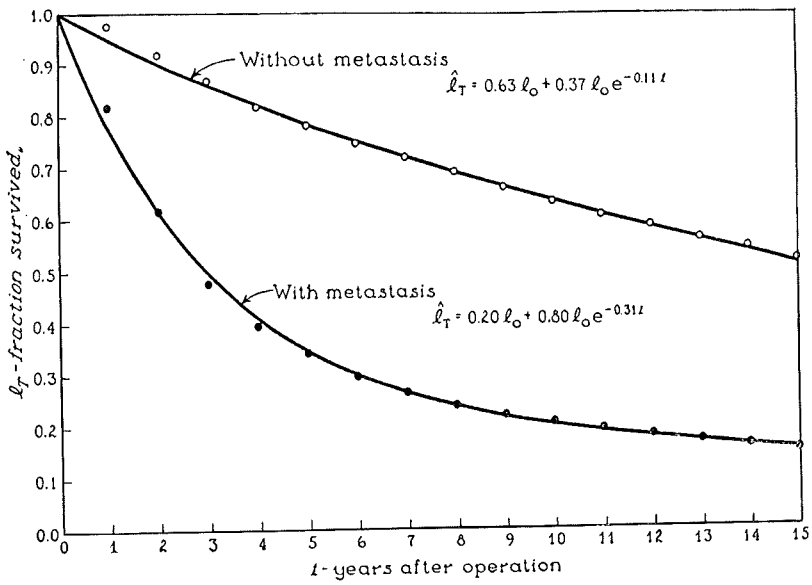


Fig. 8 - *Survival Curves for Cancer of the Breast treated by Radical Mastectomy at the Mayo Clinic.*

Reference: Berkson et al (1952).

does seem to fit the facts very well for a large sample of breast cancers.

Haybittle's discovery that the cancer death-rate falls exponentially also leads to a survival curve with two parameters. Figure 9 shows a survival curve fitted by the Method of Maximum Likelihood to some data for carcinoma of the breast treated at Manchester.

The limits of validity of these different survival functions have yet to be established. Indeed it will need very large groups of patients to reveal significant differences between the various curves. Boag has pointed out (Table 2) that if one calculates

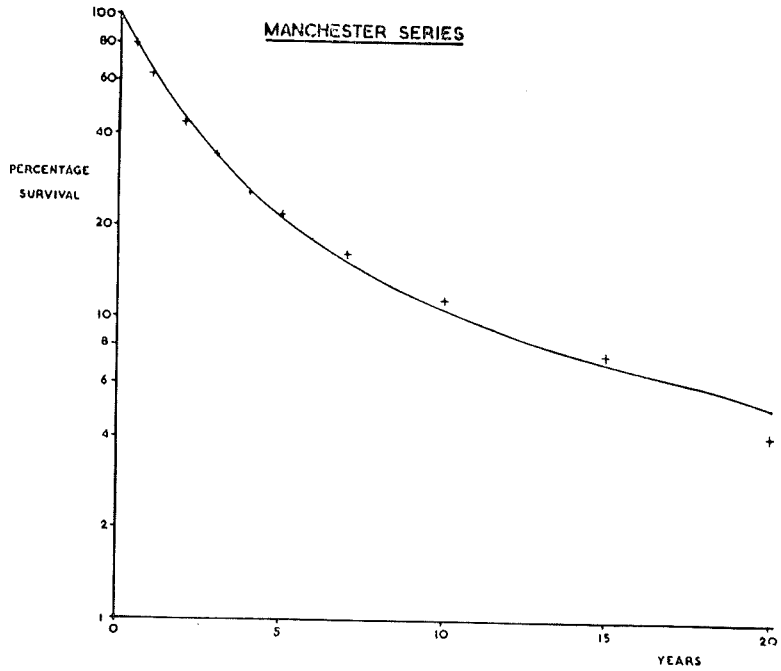


Fig. 9 - *Survival Curve for Cancer of the Breast calculated according to Haybittle's « Extrapolated Actuarial » Method.*

Reference: Haybittle (1959).

TABLE 2

ADD BENEATH TABLE: THE TABLE SHOWS THAT THE ESTIMATED SURVIVAL-RATE DOES NOT DEPEND VERY MUCH ON THE ASSUMED FORM OF THE SURVIVAL FUNCTION

M O D E L	5 *years*	10 *years*	15 *years*
Lognormal σ = 0.40	0.380	0.310	0.295
= 0.44	0.395	0.304	0.282
= 0.48	0.397	0.299	0.270
= 0.52	0.410	0.308	0.276
= 0.56	0.413	0.305	0.267
Extrapolated actuarial	0.412	0.322	0.300
Exponential	0.408	0.325	0.313

REFERENCE: Boag (1960).

relatively long-term survival-rates (as distinct from cure-rates) then it does not greatly matter which of the three formulae is assumed. This conclusion is doubtless true a fortiori if one is concerned with earlier survival-rates - say at 3 years or 5 years.

Theoretically it can be said that these parametric methods have definite advantages - provided that the assumed form of the survival curve is valid. With this proviso the Method of Maximum Likelihood gives the best possible precision in the estimation of P(t) and therefore allows conclusions to be drawn with the minimum of clinical material. However it may be that one has no previous information on the form of the survival curve, in which case one must rely upon *non-parametric methods*. I shall illustrate by examples that some of these procedures have statistical efficiencies not much inferior to their parametric counter-parts and of course their freedom from assumptions promotes added confidence in conclusions based on them.

Non-parametric Methods

The most widely used non-parametrical procedure is the Actuarial Method for estimation of survival-rates due to Greenwood. For this purpose the results are tabulated in a so-called Life-Table, as shown in Table 3. The time-scale is divided into intervals (usually years) and the conditional survival-rate for each interval is determined separately, taking account of the number of persons « exposed to risk ». If for example a patient has only been under observation for 6 months his exposure to the risk of dying of cancer is only ½; the same would apply to a patient who died of « other causes ». The overall survival-rate is the product of the conditional survival-rates for the preceding years and we can therefore establish a set of discrete points on the survival curve. The earlier points will be relatively precise, but later points will be more and more imprecise, because based on a declining amount of information. This is of course an inevitable feature of non-parametric methods, which provide no logical basis for extrapolation of the early results to longer terms.

An elegant generalisation of the Actuarial Method has been described by Kaplan and Meier. Their theory avoids the arbitrary division of the time-scale into annual increments and proves that the Maximum Likelihood estimate of the com-

TABLE 3

CALCULATION OF SURVIVAL-RATES BY THE ACTUARIAL METHOD

1	*2*	*3*	*4*	*5*	*6*	*7*
Interval following hospital dismissal, years	*Last report*		*Total persons living at beginning of interval*	*Persons, adjusted*	*Probability of dying in interval*	*Survival rate, per cent*
	Dead	*Living*				
0-1	184	—	728	728.0	0.2527	100
1-2	156	13	544	537.5	0.2902	74.7
2-3	89	8	375	371.0	0.2399	53.0
3-4	36	4	278	276.0	0.1304	40.3
4-5	31	4	238	236.0	0.1314	35.1
5-6	16	9	203	198.5	0.0806	30.5
6-7	8	3	178	176.5	0.0453	28.0
7-8	7	13	167	160.5	0.0436	26.7
8-9	6	5	147	144.5	0.0415	25.6
9-10	7	8	136	132.0	0.0530	24.5
10+	52	69	121	—	—	23.2

REFERENCE: Berkson et al. (1958).

plete survival curve is then a step-wise curve as shown in Figure 10. This diagram gives results for carcinoma of the bronchus treated with X-rays with or without the administration of cortisone.

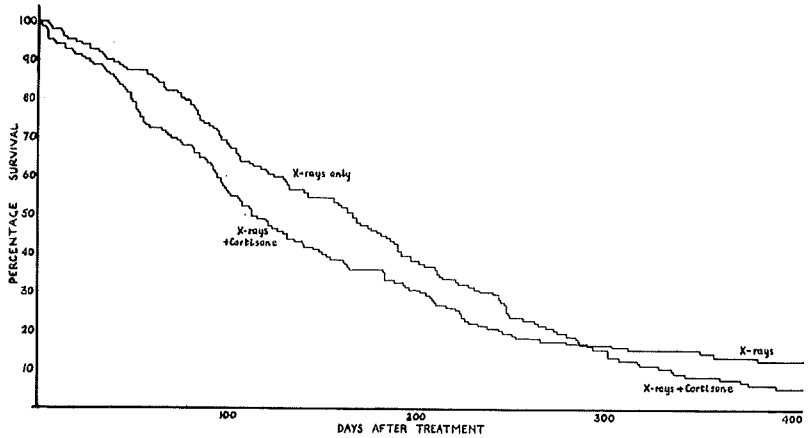


Fig. 10 - *Survival Curves of Bronchogenic Carcinoma calculated by the Product-Limit Method. Note that cortisone reduces the survival-rate - a small but significant effect.*

The same Product - Limit Method can be used to derive both the components of the survival curve. Figures 11 and 12 illustrate such an application to the study of the effect of dose-rate on radiation leukaemogenesis in mice. Figure 11 shows the survival curve analysed into its separate components due to leukaemia and other causes of death. Figure 12

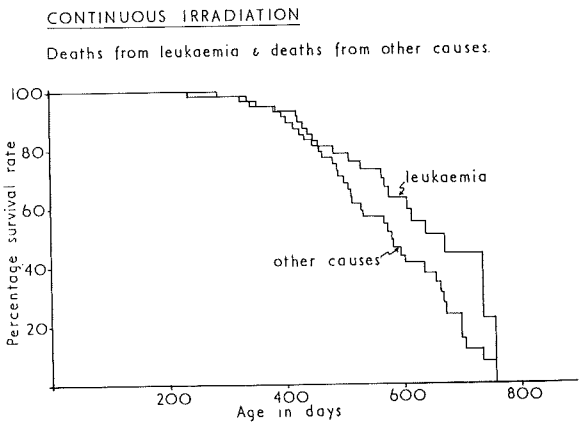


Fig. 11 - *Survival Curves for Mice irradiated with Cs. 137 Gamma-rays.*

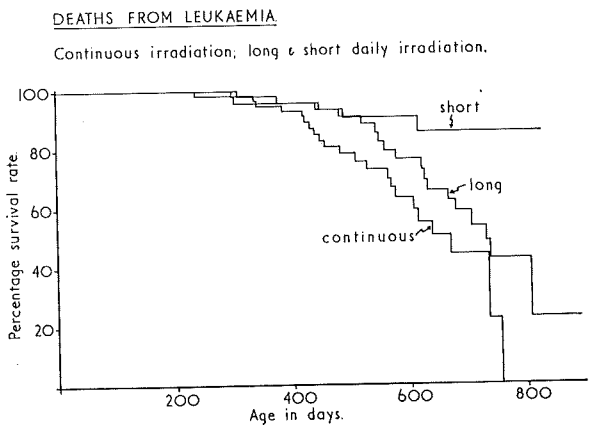


Fig. 12 - *Deaths from leukaemia in Mice Irradiated at Various Dose-Rates. Mice were irradiated to a constant dose of 1050 rads. The incidence of laukaemia is markedly higher when the dose is given at low dose-rates.*

gives the leukaemia curves for three different dose-rates, the overall dosage being 150 rad per week continued for seven weeks. It is noteworthy that deaths from leukaemia are markedly fewer when the radiation is given in short bursts rather than continuously. In view of this rather surprising result it is worth emphasizing that the method of analysis involves absolutely no a priori assumptions as to the form of the survival curves. There can therefore be no suspicion that the conclusion is an artefact introduced by the method of analysis.

Clinical Trials

So far I have been speaking of the estimation of the survival probability P(t) for a definite method of treatment applied to a defined group of cases. Of course what the physician is really interested in is how to maximise P(t). In practical terms he wishes to compare the results of two or more methods of treatment. If this comparison is done in an organised manner it is accorded the status-symbol of a Clinical Trial. The special feature about such a trial is that patients are allotted randomly to the various treatments. This is important simply because no two patients are alike and therefore we cannot divide our clinical material into precisely equivalent groups. However randomisation does ensure that the groups are precisely comparable in terms of probabilities and it thereby validates all the subsequent calculations. Without randomisation such calculations become at best doubtful and at worst completely phoney. I can illustrate the point best by examples.

Table 4 gives our earlier data for carcinoma of the bronchus, but broken down according to the time of year when the treatment was given. Notice that, in conformity with our overall finding, the cortisone group always shows the poorer result whatever the season. But suppose we had planned our experiment in such a way that patients treated with cortisone during the spring and summer had been compared with control patients treated during the previous autumn and winter. The overall result of this clinical trial would have been reversed — we might have concluded that cortisone, given in conjunction with X-rays, was of some slight benefit to the patient — a conclusion which we believe to be quite false. There must be many such pitfalls in any clinical experiment, because of unknown factors over which the clinician has no control. Randomisation automatically side-steps such hazards.

My next example (Table 5) gives results of treatment of breast carcinoma at Edinburgh and at the Mayo Clinic. The

TABLE 4

Clinical Trial of Cortisone as Adjuvant to X-ray Therapy

CARCINOMA OF BRONCHUS: 100-DAY SURVIVAL RATES

SEASON	*X-rays*	*Cortisone*
Spring	24/23 = 71%	25/39 = 64%
Summer	60/82 = 73%	53/85 = 62%
Autumn	25/35 = 71%	13/33 = 39%
Winter	30/60 = 50%	23/50 = 46%

Test for Effect of Season on Survival.

$$x^2 = 17.45,\ 6\ \text{d.f.},\ \mathbf{P = 0.009.}$$

The table demonstrates that in addition to the adverse cortisone effect there is also a « seasonal effect » i.e. survival-rates are lower during the autumn and winter than in the spring and summer. The effect is statistically significant.

TABLE 5

CARCINOMA OF BREAST: COMPARISON OF 5-YEAR SURVIVAL RATES AT EDINBURGH AND AT THE MAYO CLINIC

	Edinburgh	*Mayo Clinic*
Alive	774 (48%)	1127 (59%)
Dead	835	773
Total	1609	1900

Patients with distal metastases excluded.

Test for Difference in Survival Rates.

$$x^2 = 43.7,\ 1\ \text{d.f.},\ \mathbf{P > 10^{-6}}.$$

REFERENCES: *McWhirter*, Brit. J. Radiol., 1955, *28*, 128.
Berkson et al, Proc. Mayo Clinic, 1957, *32*, 645.

5-year survival rates are respectively 48% and 59% - an interesting difference. Let us make a statistical test - in other words let us calculate the probability of getting a difference at least as large as that observed, assuming that the effect is due to pure chance. Strictly speaking we cannot make this calculation unless we know that the two parent populations from which the patients were drawn were identical. Let us assume that they were. Then the probability in question works out as less than one in a million. It seems reasonable therefore to conclude that the difference in results was not due to chance but to the superior treatment given at the Mayo Clinic. The workers at Rochester certainly draw this conclusion, but I find that other people tend to have their doubts. Indeed the conclusion, though interesting and suggestive does not carry conviction simply because it was not possible to randomise the patients between the two treatment schemes. The results are therefore open to the criticism that the two groups of patients were not really comparable. I believe that at Edinburgh they still treat breast cancer mainly by simple mastectomy plus X-rays and at Rochester mainly by radical mastectomy. The question as to which is the better form of treatment is still an open one.

In comparing results from different centres it is often worth while to make such a provisional statistical test. If the probability works out high — say, 1 in 5 — that chance alone could have produced the given results then it is wise to reserve judgement. In the past much intellectual and emotional energy has been dissipated in the discussion of effects which were not statistically significant. It seems a pity to take pure chance quite so seriously!

Some statistical Tests

There are occasions when the conventional 5-year survival-rate is inappropriate or inefficient as an index of comparison of two or more treatments. As I have suggested already, the simplest way of estimating the survival-rate is not necessarily the best for precision, and it must be emphasized that any clinical trial will tend to give inconclusive results if the right method of analysis is not used. In other words, the finding « not statistically significant » could mean that there *was* a genuine effect, but that the method of analysis was inadequate to demonstrate it. Theory suggests that the parametric methods have advantages here. I shall illustrate the point by a few examples of parametric and non-parametric tests.

Table 6 gives the results of three tests on the cortisone and X-ray data which I have mentioned. First we have the *median test* 3-year. The straightforward comparison of 1-year, or 3-year, or 5-year survival rates is somewhat arbitrary - our survival curves might for instance diverge at the chosen time but lie fortuitously close together at other times. The median test in effect removes this arbitrariness by making comparisons at the median survival time i.e. the time at which the overall survival-rate is 50%. Otherwise the median test resembles the 5-year test in being of the simple binomial type. As you will see from the diagram the difference between the two series is only just significant on this test (P = .049).

On the other hand the *t-test,* which is appropriate on the assumption of longnormality, gives P = .016 - a more highly significant result. The diagram also gives the result of the *Wilcoxon test.* For this test the actual survival times are replaced by their ranks i.e. one simply arranges the survival times for individual patients in increasing order and assigns to

TABLE 6

Clinical Trial of Cortisone as Adjuvant to X-ray Therapy.

CARCINOMA OF BRONCHUS

1. *Median Test*

	X-rays only	*X-rays + Cortisone*	*Combined*
Below median	94	113	207
Above median	115	92	207
Total	209	205	414

$x^2 = 3.87$, 1 d.f., **P = 0.049.**

2. *Student's t-test based on Lognormal Distribution.*

t = 2.43, 416 d.f., **P = 0.016.**

3. *Wilcoxon Test.*

X-rays only, mean rank = 222.8
X + cortisone, mean rank = 195.9

$$\frac{W}{\text{dev}(W)} = 2.28, \qquad \mathbf{P = 0.023.}$$

the first the rank 1, to the second the rank 2 and so on. One then compares the mean ranks of the two groups. It is known that this simple test is efficient if the variables are normally distributed and it is interesting that the P-value is fairly similar to that given by the t-test - .023 as compared with .016. However the Wilcoxon test gives a valid result whatever the forms of the survival curves.

The next two diagrams illustrate a *non-parametric test due to Gilbert* which is closely related to Wilcoxon's test, but has the great virtue of being applicable to incomplete results. In this respect Gilbert's test bears the same relationship to Wilcoxon's as Boag's method bears to the straightforward longnormal analyses. We might therefore expect that with incomplete data Gilbert's test should be comparable with tests based on Boag's method of analysis. Figure 13 gives survival curves for two groups with carcinoma of the larynx treated during successive 5-year periods. The particular interest of this comparison is that patients in the first series were treated chiefly by conventional 250 kV X-rays, whereas those in the second series received chiefly Co 60 therapy. If we compare the two groups in terms of Boag's parameter C we see that the difference is not statistically significant (Table 7). On the other hand Gilbert's test gives P = .028, so there can be little doubt that there is a genuine difference between the two groups.

These few examples illustrate that tests based on the parametric methods are the most sensitive, provided that we do not ask too much of them — such as, for example, testing for long-term differences on the basis of short-term results. However one must immediately qualify this by saying that non-parametric tests based on ranking methods may be only slightly inferior. Binomial-type tests based on the direct estimation of proportions are definitely less sensitive and therefore require more clinical material in order to demonstrate positive results.

Concluding Remarks

In conclusion may I express the hope that I have not wearied my audience with too much statistical jargon. What I hope I have achieved is to convince you that even a simple exercise such as assessing the results of treatment involves us in fields of study far removed from traditional medical practice. Since this is an age of specialisation it would seem that a complete medical service requires the active participation of specialists in these other fields, such as physics and

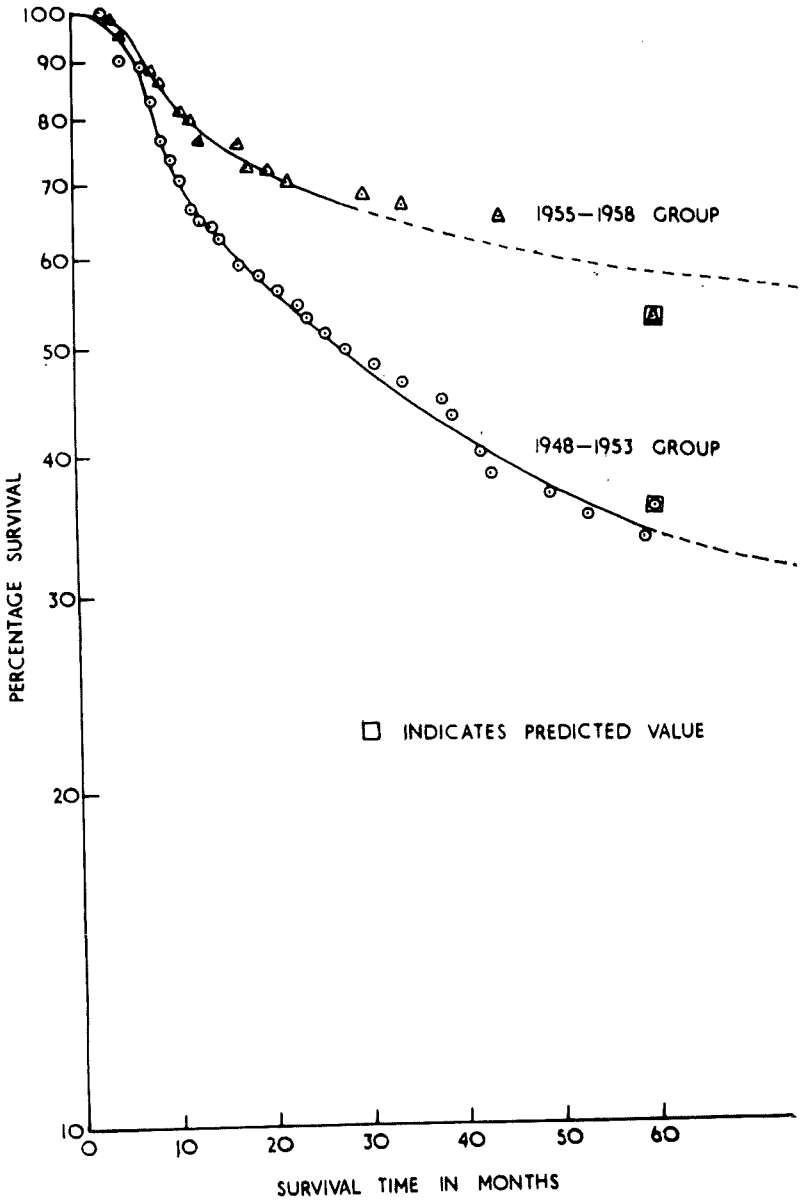


Fig. 13 - *Survival Curves for Two Troups of Patients with Carcinoma of the Larynx. Curves calculated by Boag's Method.*

Reference: Tudway et al (1960).

mathematics. I hope I have not given you the impression that the physicist would like to take on the job of the statistician as well as his own. On the contrary, what we physicists can do is to point to sections of our work which seem to demand a more quantitative approach. Dosimetry is one such field, the statistical assessment of results is another. If future developments bring about the introduction of more physicists and statisticians into the medical service then I think we physicists can see plenty of opportunities for close collaboration. On the whole I think we can conclude with some justification that experience in radiotherapy augurs well for the introduction of the more exact sciences into medical practice.

TABLE 7

Statistical Tests on Data of Tudway and Freundlich.

CARCINOMA OF LARYNX.

The clinical material consists of cases treated during 1948-53 mainly by conventional 250 kV X-ray therapy and also cases treated 1955-58 mainly by Co 60 gamma-rays. Individual survival histories are given in the publication referred to above.

1. *Test based on Boag's Parameter C.*

$C_1 = 0.334 \pm 0.059.$ $C_2 = 0.500 \pm 0.064.$

$C_2 - C_1 = 0.166 \pm 0.087$ (Binomial variances)

$u = 1.91$, **P = 0.056.**

2. *Gilbert's Non-parametric Test.*

$W = 0.187 \pm 0.085$

$u = 2.20$, **P = 0.028.**

The higher survival rate in the later series appears to be statistically significant.

REFERENCE: Brit. J. Radiol., 1960, *33*, 98.

APPENDIX

FORMULA FOR THE VARIANCE OF $\bar{P}(t)$ WHEN P(t) DIFFERS FROM ONE PATIENT TO ANOTHER

Suppose we have N patients in which the probability P(t) takes the volume P, for N, cases, P_2 for N_2 cases and so on. Define var

$$(P) = \frac{\Sigma (\bar{P}_i - P)^2}{N}.$$

We wish to estimate the mean value of P(t) i.e.

Method A. - All Cases Pooled.

If we estimate P by pooling the data ignoring the variations in P(t) then

$$\text{war } (\bar{P}) = \frac{\bar{P}\,(1 - \bar{P})}{N}$$

Method B. - Different Categories Considered Separately.

Suppose we estimate P_1, P_2 etc. separately and then calculate the mean $\bar{P}'$. It can be shown that

$$\text{var } (\bar{P}') = \frac{\bar{P}\,(1 - \bar{P}) - \text{var } (P)}{N}$$

Numerical Example.

Suppose the whole group is made up of three equal homogeneous groups.

e.g. N = N/3, P = 0.7
N = N/3, P = 0.5 Thenvar (P) = 0.027
N = N/3, P = 0.3

Method A.	*Method B.*
var $(\bar{P})$ = 0.25/N	var $(\bar{P}')$ = 0.223/N.

Thus, despite wide variations in P, the simple method using pooled data gives a reasonably precise estimate of $\bar{P}$.

REFERENCES

Berkson J., and Gage R. P., J. Amer. Stat. Ass. 1952, *47*, 501.

Berkson J., Harrington S. W., Clagett O. T., Kirklin J. W., Dockerty M. B. and *McDonald J. R.*, Proc. Mayo Clinic, 1947, *32*, 645.

Berkson J. and *Gage R. P.*, Treatment of Cancer and Allied Diseases, Vol. I (1958), p. 578. Eds.: Pack, G.T. and Ariel, I. M.

Boag J. W., J Roy, Stat. Soc., 1949, B, *11*, 15.

Boag J. W. and *Wood C.A.P.*, M.R.C. Report No. 267 (1950).

Boag J. W., Clinical Radiology, 1960, *11*, 150.

Boag J. W., Brit. J. Radiol., 1960, *33*, 199.

Fisher R. A., Phil. Trans. Roy. Soc., 1922, A *222*, 309.

Gilbert J. P., Ph. D. Thesis, Univ. of Chicago, June 1962.

Gompertz B., Phil. Trans. Roy. Soc., 1825, *115*, (1), 513.

Greenwood M., Rep. Pub. Health, No. 33 (1926).

Haybittle J. L., Brit. J. Radiol. 1959, *32*, 725.

Kaplan E. L. and *Meier J.*, J. Amer. Stat. Ass., 1958, *53*, 457.

McWhirter R., Brit. J. Radiol., 1955, *28*, 128.

McWhirter R., Clinical Radiol., 1960, *11*, 144.

Sorensen B., Acta Radiol., 1958, Suppl. No. 169.

Tudway R. C. and *Freundlich H. F.*, Brit. J. Radiol., 1960, *33*, 98.

Wilcoxon F., Biometrics, 1945, *1*, 80.

INTERVENTI SULLA RELAZIONE

J. F. FOWLER

Prof. Day, in his witty, wise and interesting paper, gave several examples of different statistical tests used to compare two sets of data. In each case some tests gave a clear indication of significance which others did not. The implication was that the tests which failed to show significance were insensitive or wrong, and that there was a significance which the better tests succeeded in revealing. Is this correct? Or can false positive results be obtained by some tests, showing more significance than is really present?

G. B. SALVATORI

Mi compiaccio vivamente col Prof. Day per il suo brillante e denso rapporto, che ha attirata la nostra attenzione, e chiarito il nostro pensiero, su metodi di rilevazione statistica nel campo della radioterapia.

Che la statistica, modernamente intensa nei suoi sviluppi squisitamente matematici debba entrare nei nostri Istituti con ampio diritto di cittadinanza, è ormai fuori discussione.

I radioterapisti come egli ha accennato, sono stati fin dall'inizio compilatori ed elaboratori di statistiche, sia pur con procedimenti elementari, ma con risultati non disprezzabili, quando hanno posto, in base ai loro calcoli, principi di importanza generale e fondamentale, che hanno incanalato ogni attività in Radioterapia e che anche attualmente ne costituiscono l'ossatura. Ad esempio: malattie da non trattare, malattie con moderata e notevole probabilità di guarigione, altre suscettibili di questa. Inoltre, metastasi trattabili o non trattabili, sopravvivenza e così via con un elenco che porterebbe troppo lontano.

Errori ne sono stati commessi, ma non bisogna credere, e gli stessi cultori di statistica confermano, che anche questa disciplina non abbia le sue trappole per usare un vocabolo usato in questo Convegno.

Circa l'argomento specifico della sopravvivenza degli ammalati di tumori maligni sottoposti a Radioterapia, osservo che la difficoltà sta relativamente nella elaborazione dei dati statistici.

Il vero scoglio è costituito dalla raccolta e dalla cernita del materiale da parte del medico.

I malati di cancro vengono a noi nelle più svariate condizioni generali e locali, alcune note, molte ignote.

Ogni tumore maligno differisce dagli altri nella sua consistenza, e quindi nel suo decorso, nell'epoca dell'inizio, nella diffusione regionale, nella diffusione metastatica, che spesso esiste, anche se sconosciuta.

Sono tutti elementi determinanti della sopravvivenza, ma che ci sfuggono.

Lo scopo del mio intervento, tralasciando altri rilievi, che, numerosi si potrebbero fare, è quello di richiamare l'attenzione su quelle condizioni che debbono formare la base di uno studio statistico sulla sopravvivenza, e facendo un passo avanti, domandare agli esperti di statistica se sia possibile applicare il calcolo delle probabilità all'insieme delle cause di errore, che io ho citato e determinarne l'incidenza sulla solidità del giudizio statistico circa la sopravvivenza degli infermi affetti da tumori maligni sottoposti a radioterapia.

R. MILANESI

Il Dott. Day ci ha mostrato l'importanza della statistica applicata alla valutazione dei risultati che si possono ottenere con le varie modalità di trattamento.

Egli ha detto giustamente all'inizio della sua relazione che la parte più importante della statistica non è quello di valutare a posteriori la bontà di un trattamento rispetto ad un altro ma bensì quello di fornire dei dati che possono in qualche modo indicare quale dovrebbe essere il tipo di trattamento da adottare per ottenere i migliori risultati.

Se non vado errato in un lavoro di Boag apparso intorno al 1957 si faceva vedere come i risultati del trattamento del tumore laringeo con il Co-60, sarebbero stati superiori a quelli che si potevano ottenere trattando gli stessi pazienti con la Rx. tradizionale. Uno dei diagrammi presentati dal dott. Day ha messo in evidenza che quella previsione era vera. Ora vorrei chiedere al Dr. Day, se la statistica attualmente è in grado o sarebbe in grado di fornire indicazioni su quali metodiche di trattamento sarebbe più opportuno adottare sia in riferimento alla scelta del tipo di lunghezza d'onda da utilizzare per effettuare il trattamento e sia poi successivamente circa la modalità tecnica da adottare per effettuare il trattamento stesso.

RISPOSTA DEL RELATORE

M. J. DAY

Reply to remarks of Professor J. F. Fowler

The degree of protection against arriving at a falsely positive conclusion is given by the P-value. Conventionally P-values under 0.05 are regarded as establishing statistical significance. The probability of reaching a positive conclusion from a statistical test, when this is in

fact the correct conclusion, is known as the « power » of the test. Tests vary in their power and a poor test may fail to give a conclusive result simply because its power is inherently low.

Reply to remarks of Dr. G. B. Salvatori

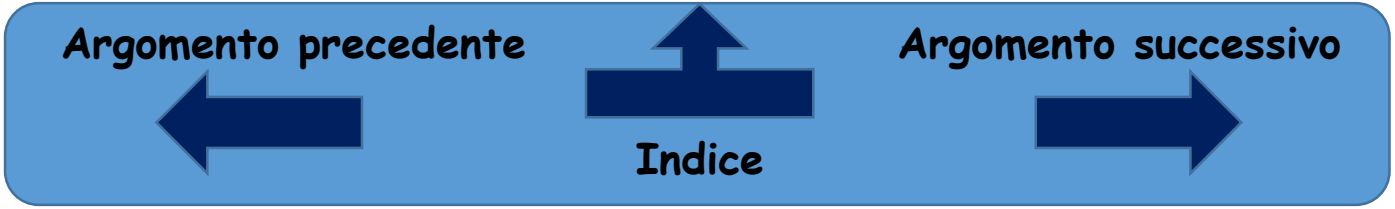
I agree with Dr. Salvatori's remarks, especially his comment that the practical difficulty is often in the collection and organisation of reliable data rather than in the mathematical analysis. The latter may be quite simple, though choice of the correct method is important.

Of course cancer is a disease which does vary widely not only in stage, pathology and other well-known factors but also in other completely unknown factors which must be postulated to explain why the survival times in two apparently almost identical patients may be so very different. These considerations should not be taken as arguments against the use of statistical methods — on the contrary it is precisely because of variations that the probabilistic approach becomes essential.

Reply to remarks by Dr. Milanesi

A statistical experiment can be designed to compare the results of any two or more different treatments. The difference may be one of technique, dosage, fractionation or of general policy (e. surgery, radiotherapy or chemotherapy, separately or in combination). I have stressed the importance of proper randomisation and have quoted a few examples of clinical trials which fulfill this requirement. In the future we can expect results of many more such experiments which re-examine scientifically questions such as those raised by Dr. Milanesi.

(PAGINA VUOTA NEL TESTO ORIGINALE)


Argomento precedente
Indice
Argomento successivo