LXI. THE DETERMINATION OF VITAMIN D.

II. A COMPARISON OF RADIOGRAPHY AND BONE ANALYSIS IN THE ESTIMATION OF VITAMIN D.

BY ROBERT BENEDICT BOURDILLON
AND HILDA MARGARET BRUCE.

From the National Institute for Medical Research, London, N.W. 3.

(Received February 29th, 1932.)

THROUGH the friendly co-operation of Miss Hume and Miss Gaffikin, and of Dr Zilva, Miss Crawford and Miss Perry, we have been afforded the opportunity of radiographing the 180 rats referred to in the preceding paper (this *Journal*, p. 488) and 200 other rats, of which the bones were analysed by Dr Zilva and his co-workers at the Lister Institute. Sufficient data are thus available for a rough comparison of the radiographic and analytical methods, and for a study of the errors in both methods.

TECHNIQUE.

Analytical method. The diet and analysis of Miss Hume's rats have been described in the preceding paper. The treatment of Dr Zilva's rats was on similar lines, except that the doses of vitamin were given in the form of natural butters or oils, of which the activity was being tested for purposes unconnected with this paper. Dr Zilva's results are expressed as the A/R ratio [Chick *et al.*, 1926], *i.e.* the ratio: weight of ash/(weight of dried fat-free bone *minus* weight of ash). The results of Miss Hume's analyses are here expressed as the percentages of ash in the dried extracted bones.

Radiography. The animals were killed by coal gas, and one knee-joint of each rat was photographed by X-rays from a hot cathode tube (dental type), using 4 amp. filament current, and 12 milliamper. high tension current at about 35,000 volts, from transformers. An exposure of 5 secs. was used, with a distance from plate to cathode of 16 inches.

A. RADIOGRAPHIC METHOD (WITH PROPHYLACTIC DOSING).

Construction of radiographic scale.

A scale for prophylactic dosing was constructed on lines similar to those previously used for a curative radiographic scale. This was described in a report [Bourdillon *et al.*, 1931], which also contained full details of the methods of estimating errors that are used in this paper. A curative scale has also been

Enlarged $\times 2$.

Natural size.

Radiographic scale for prophylactic dosing.

described by Everse and van Niekerk [1931]. For the prophylactic scale a series of photographs of developing rickets was selected, from rats used for testing vitamin D, such that each photograph showed an easily visible difference from the preceding member of the series. These are shown in Plate I, both natural size and enlarged to two diameters. It was found advisable to select only 8 stages, instead of the 12 that were used in the curative scale, since fine distinctions proved unreliable in the lower half of the scale. As with the curative scale the photographs were selected without reference to dose, and there is therefore no reason to expect any uniform relation between dose and resulting scale number. When matching photographs it is useful to recall the following criteria, which were used in distinguishing the various numbers.

Scale No. 8. Healthy bone, with no appreciable blurring of the top of the metaphysis.

Scale No. 7. Definite blurring at top of metaphysis, but only for a very short length.

Scale No. 6. Top of metaphysis blurred for appreciable distance (usually a small gap of uncalcified tissue between shaft and epiphysis).

Scale No. 5. Blurring extends to 1 mm. or more, and the bone has no definite edge to it. In some specimens the blurred area is less deep, but there is a very narrow gap of uncalcified tissue across the bone.

Scale No. 4. Definite narrow gap of uncalcified tissue right across the bone.

Scale No. 3. Uncalcified zone about 1 mm. wide.

Scale No. 2. Uncalcified zone wider, but still traces of calcification at lowest portion of new tissue.

Scale No. 1. Full rickets. Gap 1.5 mm. or more, without visible calcification.

In the earlier stages of this work a rough "verbal" scale was constructed, based on criteria similar to the above, but distinguishing 12 different stages, and used without scale photographs. Experience showed that variations in the rate of growth rendered the distinction of fine differences in the lower half of this scale rather unsatisfactory, and hence the number of divisions was reduced to 8 when making the photographic scale. However, the errors of result found with the rather crude "verbal" scale were of the same order as those of the photographic scale, and some of the results obtained by its use are quoted below. They show that it is possible to get very fair results without a photographic scale, but we do not recommend this method, as the subjective element is much larger than that involved in matching a photographic scale.

Relation between dose and effect.

Table I shows the results obtained by applying this scale to Miss Hume's rats. The litters are arranged in order of the mean ash percentage of the litter after dosing, beginning with the lowest. These data are illustrated in Fig. 1, which shows the curves obtained by plotting the logarithm of the dose

of vitamin against the radiographic scale number. The central curve No. 2 shows the mean results, for all litters, each point representing 20 rats. Curves 1 and 3 show the results for the groups of 7 litters that showed the highest and lowest ash percentages respectively.

Table I. *Radiographic scale readings of rats' knee-joints after prophylactic dosing.*

Litter No.	Dose (in International Units)									Mean ash percentage for each litter for doses 0 to 0.32 only
	0	0.01	0.03	0.05	0.1	0.2	0.32	0.64	1.28	
	Scale readings									
1867	1	3	1	1	3.5	5	5	—	—	31.36
1869	2.5	4	2.5	2.5	3.5	5.5	7	—	—	35.66
1828	4	4	4	6.5	6	6	8	—	—	37.01
2057	2	1	3	5	5	6	8	7.5	7.5	37.26
1834	5	5.5	4.5	5.5	6	7	7	—	—	38.39
2060	2	1.5	5	5	5	7	8	8	8	38.87
1827	5	4	7	5	6	6	8	—	—	38.97
Mean of 7 lowest litters	3.07	3.29	3.86	4.36	5.0	6.07	7.29	—	—	36.79
1854	2	4	4	5	6	7	6	—	—	39.76
1875	2	5	5	5.5	5.5	8	6	—	—	40.53
1914	5	5	5	5.5	6	6	7	8	8	41.84
1826	5	4	5.5	5	6	6	8	—	—	41.97
2095	1	2	3.5	4	4	5.5	6.5	8	7	42.14
1991	4	5	6	5.5	5	5.5	7.5	7.5	7.5	42.74
1915	4	7	4.5	5	6	7	6	7	7.5	43.01
1917	5	4	4	5.5	6	6	8	7	7.5	43.11
1988	3	5	6	5	5	5.5	7	6.5	7	43.13
1965	5	5	6	5	6	5.5	7	8	8	43.56
1943	5	5	5	4.5	5	6	7	8	7.5	44.06
1963	5	5	5	5	6	8	8	8	8	44.66
1954	5	5	6	6	7.5	7	7	8	7	48.61
Mean of 7 highest litters	4.57	5.14	5.21	5.14	5.93	6.45	7.14	7.5	7.5	44.31
Mean for all litters	3.62	4.2	4.62	4.85	5.42	6.27	7.1	7.62	7.54	40.83

These curves diverge with decreasing dose to an important extent. For doses of less than 0.05 (International) unit the highest litters give a nearly horizontal curve, suggesting either that serious amounts of vitamin were present in the diet or stored in the animals' bodies, or else that some other factor influenced conditions in such a way as to render the development of full rickets impossible. This point is discussed later.

The convergence and approach to parallelism of the upper portion of the three curves is noteworthy (especially when contrasted with the results of bone analysis shown in Fig. 2). It suggests that the dose of vitamin required to prevent the development of faint traces of rickets is much more constant than the doses required to prevent the development of full rickets, and may be more constant than those needed for the cure of developed rickets, in different litters. If this were the case, it would account for the statements of certain

workers, who use as their criterion "complete protection" from radiographic rickets, instead of cure or bone analysis, that they can obtain a stock of practically constant sensitivity without great difficulty.

Fig. 1. Radiographic scale. Relation between dose and effect¹.

Evaluation of results from radiographic scale.

Owing to the curvature of the centre line in Fig. 1, it is more satisfactory to evaluate results by graphical methods than by an equation. If rats are dosed in pairs of litter-mates on standard and unknown solutions, it would be strictly correct to evaluate each pair separately. But in practice, if the doses used do not cover a wide range, it is sufficiently accurate, and much less laborious, to use the mean scale reading of the rats on standard solution, and the mean reading for the rats on corresponding doses of the test solution. The logarithms of the doses corresponding to these two mean values are then read off from the curve, and the antilogarithm of the difference between the two logarithms equals the ratio of the strengths of the standard and unknown solutions.

Since, however, the slope of the curve varies appreciably with the sensitivity of the stock, it is not desirable to use one curve for all tests, unless the stock is fairly constant; but for doses above 0.05 unit, for which alone this method is suitable, this variation in slope is not very great, and the plotting

¹ For convenience the unit of dose used when plotting logarithms was 1×10^{-6} mg. of irradiated ergosterol. This is one-hundredth of the International unit which is used elsewhere throughout this paper.

of a separate curve for each test is only necessary when the doses of standard and unknown solutions have been chosen so as to give mean scale readings of widely different values. Whenever possible the doses should be chosen so as to give approximately equal quantities of vitamin from the standard and the test solutions. When this has been done, the curve is only used to apply a small correction, and minor variations in slope are then of small importance.

Fig. 2. Bone analysis. Relation between dose and effect.

Errors in radiographic estimation.

Terminology. In order to compare the errors of different methods of estimation of a vitamin, it is desirable to express all errors in terms of dosage. This can be done whenever the relation between dose and effect is known. It is then convenient to express any error as the ratio "apparent dose"/"true dose," or as the percentage error in estimation of dose which would be caused by the error in question. Thus an error of +0.5 scale division in the centre portion of the radiographic scale can be seen from the curve in Fig. 1 to correspond to a rise of 0.24 in the logarithm of the dose. Since the antilogarithm of 0.24 is 1.738, this error equals a dose ratio "apparent dose"/"true dose" of 1.74, or an error of +74 %. It should be noted that a negative error, such as -1 scale division, gives a dose ratio which is the reciprocal of the value for the same positive error (e.g. 1/1.74): but when expressed as a percentage, a negative error has a smaller numerical value than a positive error. Thus an error of +100 % corresponds to a negative error of -50 %. Throughout this paper we express as percentages positive errors only, and we obtain them by subtracting

1 from the dose ratio, and multiplying the remainder by 100. We also use the letter σ to mean standard deviation.

In practical testing it is useful to know the frequency with which any given error can be expected. We have therefore calculated (a) the "probable error," which corresponds to $2/3\sigma$, and is likely to be reached or exceeded once in every two tests, (b) the error corresponding to 2σ , which is likely to be reached or exceeded once in every 22 tests. We have usually estimated these figures on the assumption that each test is based on 20 pairs of rats (half on standard and half on test solution). The corresponding errors for any other numbers of pairs, say x , can be calculated by multiplying the standard deviation for the error in question by $2\sigma/\sqrt{x}$, or $2/3\sigma \div \sqrt{x}$, and translating the product into terms of dose ratio.

The chief errors in radiographic estimation may be grouped into four classes.

- (1) Errors in estimation of degree of rickets expressed as scale numbers.
- (2) Errors in evaluation of the scale numbers in terms of dose.
- (3) Errors due to variation between different litters in response to vitamin.
- (4) Errors due to variation between litter-mates.

(1) *Errors in estimation of degree of rickets expressed as scale numbers.* These are termed the "scale-matching error." Among any batch of photographs, some will resemble the scale photographs so closely as to allow of no doubt in allotting the scale numbers to which they correspond. But other photographs will not resemble any scale photograph precisely, and variations of one-half, or even one division, may easily occur in matching. These errors were tested as follows.

One observer matched 196 photographs twice, with an interval between the observations. The following results were obtained:

Difference in scale divisions between first and second reading	Frequency of differences	Product: Error $\times$ frequency
0.0	114	0
± 0.5	66	33
± 1.0	16	16
		<hr/> 49

Average error 0.25 division; sum of + differences 20; sum of - differences 29.

The standard deviation of the differences is 0.407 scale division. This, converted into terms of dosage (using the value for doses between 0.1 and 0.2 unit), gives a probable error of 5.5 % on a test using 20 pairs of rats.

This value is high, and amounts to about half the total error observed in practice (*cp.* Table II). We suspect that the error of matching in practical work is lower than that observed in this test, since in practice the radiographs of litter-mates are matched in rapid succession, and this removes the likelihood of a change of judgment occurring during the process. With regard to the differences recorded in the above table, it should be noted that a considerable

proportion of differences of 0.5 unit and 1 unit is to be expected in matching to any scale.

2. *Errors in evaluation of the scale numbers in terms of dose.* This error becomes zero when doses of test solution are chosen so as to give results equal to those obtained by the standard solution. In practice this is rarely attained, but if the stock used for any particular test is fairly uniform, and if the mean dose ratio of test solution/standard solution is between 1.5 and 0.66, and is calculated from a curve plotted for the rats in question, the error should only be small. If, however, the test results in a dose ratio of test solution/standard > 3 or < 0.3 , we think there is a possibility of large error with any method of estimation of vitamin D.

3. *Errors due to variation in sensitivity of different litters.* Errors due to interlitter variation may be very serious, or be entirely obviated, according to the method of testing. If tests are conducted by dosing pairs of litter-mates with standard and unknown solution, interlitter variation can only cause negligible errors. If no standard solution is used, the errors may be very large, as in the example given later in this paper. A common practice is to mix 6 or 7 litters, and to divide doses of standard and unknown solutions as uniformly as possible between the various litters. This reduces the interlitter errors, but does not entirely obviate them.

4. *Errors due to intralitter variations in sensitivity.* The variation in the sensitivity of litter-mates is considerable, although much smaller than that between different litters. It is difficult to estimate directly, but the combined error of result due to this cause, and to the scale matching error, is easily estimated as follows. The scale readings are tabulated for a series of pairs of litter-mates given two different doses in a constant ratio. The standard deviation of the differences between the readings for the two doses is then calculated in scale divisions, and converted into terms of dosage by the use of a curve such as is shown in Fig. 1. The results of this procedure, when applied to the radiographic readings, are shown in Columns 2, 6 and 7 of Table II. The upper portion of the table shows the errors for Miss Hume's rats for each pair of doses. Each of these results is calculated from σ for 20 pairs of rats, except for the highest dose pair, for which only 12 pairs of rats were used. It will be noticed that the errors fall steadily, from an enormous value for the smallest doses, to a relatively small value for the highest doses. The excessive error for the small doses confirms the opinion, which would be formed from mere inspection of the figures in Table I of this paper and Table IV A of the preceding paper, that the methods are unsuitable for doses of less than 0.05 unit. There are two reasons for this. Firstly, in the lowest part of the radiographic scale a small error in scale number causes a very large error in apparent dose, and, secondly, the actual variations in response of litter-mates appear greater, whether judged by bone analysis or by radiography. This is shown by the high values of σ for small doses in Columns 2 and 4 of Table II. The range of dosage

over which the radiographic scale is reasonably satisfactory is 0.64 to 0.05 unit with the rats concerned, covering, therefore, a range of about 12:1. The range is limited at the lower end by undue errors in result, and at the higher end by the completeness of protection against rickets. Inspection of curve 3 of Fig. 1 suggests, however, that with a stock of rats showing high sensitivity to rickets, the useful dose range might increase to 24 to 1 or even more, owing to the increase in steepness and smoothness of the curve for such litters at small doses.

Table II. *Errors of result calculated from radiographic and analytical data for tests on 20 pairs of rats.*

A. Miss Hume's rats.								
Col. 1 Doses concerned (units)	Col. 2	Col. 3	Col. 4	Col. 5	Col. 6	Col. 7	Col. 8	Col. 9
	Deviations for difference between 2 doses				Errors by radio- graphy with pro- phylactic scale. % error ex- pected once		Errors by bone analysis. % error expected once	
	Radiography		Bone analysis					
	σ (in scale divs.)	Mean difference in scale divs.	σ (in ash %)	Mean difference in ash %	In 2 tests	In 22 tests	In 2 tests	In 22 tests
0.01 and 0	1.268	0.573	3.175	1.27	—	—	—	—
0.03 „ 0.01	1.477	0.425	3.847	1.25	71	401	66	255
0.05 „ 0.03	1.030	0.225	2.736	1.53	43	145	15	53
0.10 „ 0.05	0.739	0.625	2.952	2.025	15	51	16	57
0.20 „ 0.10	0.914	0.80	2.313	2.035	12	41	13	43
0.64 „ 0.32	0.7672	0.375	2.628	1.10	22	82	14	50
B. Dr Zilva's rats (26 pairs).								
Variable doses. Ratio constant at 3:1	1.541	2.312	A/R ratio 0.1036	Mean 0.1870	11.5	38.7	11.1	32
C. Curative radiographic method [Bourdillon <i>et al.</i> , 1931].								
Variable doses. Ratio constant at 2:1	1.529	2.0		(1931) 8.2	26			
	1.458	2.0		(1932) 8	25			

The central section of Table II shows the errors for 26 pairs of Dr Zilva's rats, given varying doses of butters in the ratio of 3 to 1 for each pair of litter-mates. The results agree well with those from Miss Hume's rats, and show close concordance, whether calculated from the photographic X-ray scale, or from the "verbal" scale mentioned above.

At the foot of the table corresponding figures are introduced for comparison from the curative radiographic scale. In this series each figure refers to results from 1000 pairs of rats.

It is interesting to note that Bills *et al.* [1931] calculated by a different method from that used above that the probable error of a line test using 20 rats was 10 %. If this is multiplied by $\sqrt{2}$ to give the error for differences between pairs, it gives an error of 14 %, which is of the same order as the probable errors found above.

Frequency of errors.

The statement that an error of 2σ and $2/3\sigma$ would occur once in 22 tests, and once in 2 tests respectively, rests on the assumption that the frequency of errors of different magnitude follows the "normal curve" of error. Unfortunately, the data available are too few to allow of drawing a satisfactory frequency curve showing errors of different magnitudes. We have, therefore, counted the frequency of errors of or exceeding $2/3\sigma$ and 2σ in the difference table for Miss Hume's rats, to see how far the assumptions made in Table II were justified. In 100 differences the errors of or exceeding $2/3\sigma$ were 51 for the bone analysis, and 52 for the radiography, agreeing well with the expected number of 50. The errors of or exceeding 2σ were 5 for radiography and 8 for bone analysis, the expected number being 5. There is thus an abnormal proportion of large errors in the bone analysis data, but the numbers are too small to show how far this is significant.

Frequency of errors exceeding 100 % or other given ratio. A rough estimate of the accuracy of any method can be given by noting the frequency of "reversals" for pairs of doses in any given ratio, *i.e.* of instances when the rat on the lower dose gives a higher result than a litter-mate on the higher dose. Table III shows the frequency of such cases in these results.

Table III.

Observer	Doses concerned Units	Dose ratio	Reversals per 100 pairs	
			Bone analysis	X-ray scale
Hume	0.03 and 0.01	3:1	35	20
"	0.10 " 0.05	2:1	30	10
"	0.20 " 0.10	2:1	10	10
"	0.64 " 0.32	2:1	25	25
Zilva	Various	3:1	4	4
Hume	0.20 and 0.05	4:1	10	5

For comparison it may be mentioned that the frequency of reversals over a dose ratio of 2:1 with the curative radiographic scale, at the Institute for Medical Research, was found to be 99 in 2000 pairs, *i.e.* 4.55 per 100 pairs. While of interest as a rough test, this type of frequency table is too liable to errors of sampling to be a useful guide for comparison of different methods, except when dealing with many hundreds of pairs.

B. BONE ANALYSIS.

Relation between dose and effect.

Fig. 2 shows the results obtained by plotting the ash percentage in the case of Miss Hume's rats against the logarithm of the dose. The centre curve shows the mean results for 20 litters, the upper and lower curves show the mean results for the 7 litters yielding respectively the highest and lowest calcium percentages. The graph is of interest for four reasons.

1. The two lower curves are approximately straight over a considerable range of dose (from 0.03 to 0.32 or 0.64 International Units), and the upper

curve, although less regular, has a general course which suggests that it would be linear over the same range, were it not for random errors in the small number of rats represented (7 per point).

2. The three straight lines that can be drawn through these data converge to a point, which represents a calcium percentage of 55.2 and a dose of 10 units. However, increase of dosage of vitamin beyond 0.64 unit does not cause a rise in ash percentage, but even appears to cause a slight fall, as can be seen from Table IVA of the preceding paper. This fall is not confined to the litters showing highest average calcium, but is also seen in the litters showing low average calcium, such as Nos. 2057 and 2060, although in these litters doses of 64 mg. give results much below the highest figures reached in the series. A similar fall with increasing dose was found by Schultzer [1931].

These facts strongly suggest that the variation in response to vitamin shown by different litters is caused, not by storage of vitamin, or by traces of vitamin present in the diet at different seasons, but by a second variable of different nature, possibly some other constituent of the diet. This view is supported by the dotted curve in Fig. 2. This curve shows the results of assuming that each rat in the 7 litters showing high ash percentage had contained stored vitamin equivalent to a dose of 0.1 unit per day. Thus doses recorded as 0.01, 0.03, 0.05, 0.1, 0.2 and 0.32 units are plotted as 0.11, 0.13, 0.15, 0.20, 0.30 and 0.42 units respectively. Although this brings the lowest point on the curve to the centre straight line, the rest of the curve does not follow this line at all, but approaches the upper line.

3. The results are widely different from the X-ray curves, which converge to a point within the diagram corresponding to a dose of 0.3 unit. This is clear evidence that the physiological changes detected by bone analysis and by radiography are not equally affected by the different variables which control the response to vitamin in rats.

4. The three curves lie at widely different heights above the base line, corresponding to great differences in sensitivity of the various litters. Thus the 7 litters in curve 1 gave the same ash percentage with a dose of 0.01 unit, as did the 7 litters of curve 3 after a dose 32 times as great. This shows that, in testing with a stock of this range of variation, failure to use a standard solution might cause errors of over 1000 %.

Evaluation of results of test.

The graphical method of doing this has been discussed in the preceding paper, and is convenient for most purposes. Although the slope of the lines varies too greatly to allow of the use of the simple equation given for the logarithmic curative scale in our previous report, the following general equation may be used for evaluating results when a non-graphical method is desired:

$$\text{Log } D_2/D_1 = \frac{(n_2 - n_1) \log 2}{x}$$

where D_2 and D_1 are doses which produce ash percentages of n_2 and n_1 , and x is the increase in ash percentage caused by doubling the dose of vitamin. This equation may be more conveniently expressed as Dose ratio = Antilog (ash difference $\times 0.3010/x$). In Fig. 2 the values of x are 1.4, 2.0 and 2.4 for the upper, middle and lower lines respectively.

Errors in estimation by bone analysis.

These errors are closely similar to the errors discussed above, except that the error of analysis is substituted for the scale-matching error, and that the errors due to variations between different rats may be of different magnitude in the two methods, since the physiological changes used for estimation are not identical.

1. *Error of analysis.* There is no simple method of measuring this, but an estimate can be obtained by comparing the results for right and left legs described in the preceding paper. As there explained, this result shows the sum of the error of analysis and of any true asymmetry that may be present, and can therefore only be regarded as a maximal value for the error of analysis. The standard deviation for this difference is 0.5833, expressed in terms of ash percentage. When converted into terms of dosage this gives a probable error of 3 % in tests based on 20 pairs. The error is thus appreciable, but is only about $1/5$ of the error of result shown in Table II, and about one-half of the scale-matching error found for the prophylactic radiographic scale.

The following test makes it seem probable that the asymmetry in calcification is very small, and that the chief part of the above errors is due to errors of analysis. The right and left legs of about 80 rats were compared by separate analyses and by radiography. In order to avoid any possibility of bias, the right and left legs were matched against the scale at different times, although owing to the different appearance of left legs from the right legs of which the scale is composed, this procedure involved a rather large scale-matching error. It was found that of the 7 pairs in which the ash percentage of the two legs appeared to show a difference of 1 % or more, the radiographic scale numbers were identical in six pairs, and only showed a difference of 0.5 division in the seventh. Of the 16 pairs in which the radiographic scale divisions appeared to show a difference of 1 or 1.5 divisions between the two legs, only 3 showed a difference in ash percentage larger than 0.6, and these were in the opposite direction to the difference in scale reading. Further, in the 9 cases showing a higher scale number in the left leg than in the right, the average ash percentage was lower in the left leg than in the right. This complete absence of correlation between radiography and bone analysis when applied to the differences between right and left legs, is strong evidence that the apparent differences are due to errors of analysis and of scale-matching rather than to true asymmetry. These results differ from those of Adams and McCollum [1928] who found evidence of appreciable asymmetry with bone analysis after curative dosing.

2. *Errors of result (including errors due to intralitter variation).* As with the

radiographic method, the combined effect of errors of analysis, and errors due to variation between litter-mates, can be obtained by calculating the deviations for pairs of litter-mates that have been given different doses of the same solution. The results of this procedure are given in Columns 4, 5, 8 and 9 of Table II. They indicate that the method is satisfactory for doses between 0.64 and 0.03 unit, but too erratic for smaller doses.

An alternative method of estimating errors of result is to calculate the deviations from the mean for a group of rats that have received the same dose of vitamin. In order to get a figure comparable with the standard deviation for differences between pairs the figure thus obtained must be multiplied by $\sqrt{2}$.

The series of Dr Zilva's rats contains 16 groups, each of 6 or 8 rats from a group of litters of approximately the same date, that have been fed on one dose for each group. The standard deviations of the differences for each rat from the mean for its group were calculated for each group separately from the formula $\sigma^2 = \Sigma x^2 / N - 1$, where x is the deviation for one rat from the mean for its group and N the number of rats in each group. The square roots of the mean of the squares of the 16 values of σ thus obtained were 0.089 for the A.R. ratio, and 2.286 for the X-ray readings on the "verbal scale." On multiplying by $\sqrt{2}$ to get a value for differences between pairs, and then converting into terms of dosage by using the relation between dose and effect found for the 26 pairs quoted in Table II, this gives a probable error for a test on 20 pairs of rats of 12% when calculated from the A.R. ratio, and 12.5%, if calculated from the X-ray scale. These figures slightly exceed those shown in Table II, as would be expected, since this method of calculation is less reliable than that used for Table II, the results being increased by variations in sensitivity between different litters. In the case of Dr Zilva's rats, the litters receiving any one dose were of approximately the same date, and the interlitter error was small. On the other hand, the application of this method of calculation to Miss Hume's series leads to excessive estimates of error, as would be expected, since the litters including rats on any one dose varied greatly in sensitivity at different periods of the year.

C. COMPARISON OF SOME DIFFERENT METHODS OF ESTIMATION OF VITAMIN D.

The above data afford a basis for making some comparisons between bone analysis and radiography, and, when compared with data already published, between prophylactic and curative methods. Such comparisons are most easily made under separate headings, according to the method of dosing, and of estimating the effect.

(a) *Comparison of bone analysis versus use of radiographic scale, with prophylactic dosing.* The following conclusions appear justified.

1. The accuracy of estimation of the particular effect measured is greater with bone analysis than with radiography (ratio of errors about 1 to 2), and

the small subjective element which is introduced into radiography by scale-matching is almost absent from the method of bone analysis.

2. The range of doses, and sensitivity of stock, over which satisfactory measurement of effect is possible is greater with bone analysis than with radiography (in the ratio of about 2 to 1). This is of appreciable advantage in practical testing.

3. The initial cost of (new) radiographic outfit is about £100, whereas bone analysis needs little equipment.

4. The error of result observed in tests of doses of known ratio appears approximately equal in the two methods over the more limited range of doses for which the radiographic method is suitable. This suggests that the metaphyseal changes in calcification are more strictly proportional to dose of vitamin than are the changes in the rest of the bone. The increased accuracy of analysis over scale-matching appears to be largely counteracted by this factor.

5. The chief source of error in both methods appears to be variations in the sensitivity of litter-mates. This error is reduced in proportion to the square root of the number of animals used for each test. Hence it is desirable to use as many animals as possible, and therefore the time and labour required become important factors, not merely in the cost of testing, but in the accuracy obtainable in routine tests. In this respect the speed of the radiographic estimation gives it a great advantage over bone analysis. The time required for radiography is about 2 minutes per rat, including killing and checking numbers. The scale-matching takes rather less than 1 minute per rat, including calculation of results. The development and washing of plates takes an appreciable time, but can easily be taught to unskilled boys.

(b) *Comparison of bone analysis with radiography, using curative dosing in both cases.* Owing to the irregular results obtained with bone analysis [cf. Schultzer, 1931; Hess *et al.*, 1931] when using curative dosing, there can be little doubt that radiography is preferable in this instance.

(c) *Comparison of bone analysis, using prophylaxis, versus radiography, using curative dosing.* In this comparison the following conclusions seem warranted.

1. For tests involving any given number of pairs of rats, there is not a great difference in accuracy between the two methods. The deviations observed in the prophylactic bone analysis series on 300 rats are from 50 to 100 % larger than those observed previously with the curative radiographic method but the difference is not large enough to be of certain significance on only 300 rats.

2. For routine estimation of vitamin D on any extensive scale the radiographic method is preferable, since the decreased labour involved allows of the employment of a larger number of rats, with consequent increase in accuracy.

(d) *Prophylactic dosing, radiographic method. Comparison of scale versus method based on "characteristic curve."* We have elsewhere [Bourdillon *et al.*, 1931] discussed methods based on a determination of the percentage of rats which, when given a uniform dose, reach any given standard of rickets. Such

methods are widely used on the continent [Holtz *et al.*, 1931], and are in effect based on principles similar to those discussed by Trevan [1927] and illustrated by "characteristic" or frequency curves. We concluded that, when using curative dosing, the use of a radiographic scale was preferable to the use of a "characteristic curve." However, when using prophylactic dosing the advantages of the characteristic (or frequency) curve became much more obvious. Consideration of the curve in Fig. 1 shows that the radiographic scale is most sensitive and most reliable at the upper end, *i.e.* that rats which are just developing the first trace of rickets show a more certain response to a given change in dose of vitamin than do rats with well developed rickets. Further, the range of dose over which the prophylactic scale is useful is limited. In view of this we think it an open question whether, if prophylactic dosing is used with radiography, it is better to use a scale, as above, or merely to determine the percentage of rats which, on a given dose, develop any visible trace of rickets, and to apply small corrections by a characteristic curve.

(e) *Comparison of prophylactic dosing versus curative dosing, using radiography for measurement of effect.* The prophylactic scale is inferior to the curative scale in that it covers a smaller range of dosage, and in the lower half of the scale is less accurate. The relation between effect and the logarithm of the dose is less simple, and varies at different parts of the scale. Further, variations in sensitivity of the stock affect the prophylactic scale and the relation between dose and effect more seriously than they do the curative scale. It is probably owing to these differences that the average observed error with the prophylactic scale is appreciably greater than with the curative scale.

Comparisons with results of other workers.

The relative merits of curative and prophylactic dosing have been discussed frequently. Thus Schultz [1929] and Poulsson and Lövenskiöld [1928], have stressed the advantages of curative dosing, on the grounds that rats always vary, and that radiography before dosing is useful in detecting such variation and eliminating the grosser cases. We have found this a great advantage in our own curative work. On the other hand, many workers advocate prophylactic dosing as superior to curative, and some regard rats used in this way as of such unvarying sensitivity that they can serve as a reliable basis for definition of a vitamin unit, without reference to a standard [Scheunert and Schieblich, 1929; Holtz *et al.*, 1931]. However, Miss Hume's results show that prophylactic dosing does not necessarily prevent very wide variation in sensitivity of stock, and Holtz *et al.* [1931] admit that their own stock varies in sensitivity by more than 100 %.

It seems to us that the choice of method of dosage should depend on the method of detection of results. Thus there is little doubt that bone analysis is unsatisfactory with curative dosing, as is shown by Adams and McCollum [1928] and by the careful comparison made by Schultzer [1931]. This is

probably due to the fact that when a severely rickety animal is given vitamin D in moderate doses, calcium deposition begins at the metaphysis, and only affects the diaphysis later, and more irregularly [Hess *et al.*, 1931]. On the other hand, the line test, which has been studied in detail [Bills *et al.*, 1931; Dyer, 1931], is usually considered more satisfactory with curative dosing than with prophylaxis.

When radiography is used, the advantages of the two methods of dosing are more evenly balanced. We think that some of the arguments used against curative dosing with radiography are not unanswerable. Thus Schultzer states that it takes a longer time than prophylactic dosing. We have, however, found a total period of 28 days (14 preparative and 14 on dose, using a Steenbock diet) very satisfactory, although this period is identical with that used by most workers for prophylactic work. Again, it is possible that the irregular results found by Schultzer for curative radiography may depend to some extent on the rather short curative period of 7 days used by him. We have not ourselves been able to get satisfactory results in so short a time. The curative radiographic scale has also been criticised by Holtz *et al.* [1931] on the ground that a number of stages of healing must be distinguished. This is, however, a real merit in practice, since it allows of getting useful results from tests over a considerable range of dose, with a resultant decrease in labour when testing solutions of unknown concentration.

SUMMARY.

1. A radiographic scale is described, showing various degrees of development of rickets, for use in estimation of vitamin D by measurement of its prophylactic effect.
2. The errors in estimation of the vitamin when using this scale and when using bone analysis are compared. While the error in determination of the resultant effect is considerably greater with the radiographic scale than with bone analysis, the total error of a test with any given number of rats appears roughly equal in the two methods provided that the doses of vitamin are confined to the more limited range of the radiographic method.
3. Owing to the decreased labour of the radiographic method it is easy to use a larger number of rats, and so to get greater accuracy than with bone analysis.
4. Study of the relation between dose and effect in the two methods suggests that the chief variations in the sensitivity of a stock of rats fed on a supposedly constant diet are due to some factor other than vitamin D.
5. The prophylactic radiographic scale is less satisfactory than the curative scale.

REFERENCES.

- Adams and McCollum (1928). *J. Biol. Chem.* **78**, 495.
Bills, Honeywell, Wirick and Nussmeier (1931). *J. Biol. Chem.* **90**, 619.
Bourdillon, Bruce, Fischmann and Webster (1931). *Sp. Rep. Ser. Med. Res. Coun. Lond.*, No. 158.
(H.M. Stationery Office, London.)
Chick, Korenchevsky and Roscoe (1926). *Biochem. J.* **20**, 622.
Dyer (1931). *Quart. J. Pharm.* **4**, 503.
Everse and van Niekerk (1931). *Ned. Tijdschr. Geneesk.* **75**, 1101.
Hess, Berliner and Weinstock (1931). *J. Biol. Chem.* **94**, 9.
Holtz, Laquer, Kreitmair and Moll (1931). *Biochem. Z.* **237**, 247.
Poulsen and Lövenskiöld (1928). *Biochem. J.* **22**, 135.
Scheunert and Schieblich (1929). *Biochem. Z.* **209**, 290.
Schultz (1929). *Z. Kinderhkl.* **47**, 449.
Schultzer (1931). *Biochem. J.* **25**, 1745.
Trevan (1927). *Proc. Roy. Soc. Lond. B* **101**, 483.