THE LANARKSHIRE MILK EXPERIMENT.

By "STUDENT."

In the spring of 1930 * a nutritional experiment on a very large scale was carried out in the schools of Lanarkshire.

For four months 10,000 school children received $\frac{3}{4}$ pint of milk per day, 5000 of these got raw milk and 5,000 pasteurised milk, in both cases Grade A (Tuberculin tested); another 10,000 children were selected as controls and the whole 20,000 children were weighed and their height was measured at the beginning and end of the experiment.

It need hardly be said that to carry out an experiment of this magnitude successfully requires organisation of no mean order and the whole business of distribution of milk and of measurement of growth reflects great credit on all those concerned.

It may therefore seem ungracious to be wise after the event and to suggest that had the arrangement of the experiment been slightly different the results would have carried greater weight, but what follows is written not so much in criticism of what was done in 1930 as in the hope that in any further work full advantage may be taken of the light which may be thrown on the best methods of arrangement by the defects as well as by the merits of the Lanarkshire experiment.

The 20,000 children were chosen in 67 schools, not more than 400 nor less than 200 being chosen in any one school, and of these half were assigned as "feeders" and half as "controls," some schools were provided with raw milk and the others with pasteurised milk, no school getting both.

This was probably necessary for administrative reasons, owing to the difficulty of being sure that each of as many as 200 children gets the right kind of milk every day if there were a possibility of their getting either of the two. Nevertheless, as I shall point out later, this does introduce the possibility that the raw and pasteurised milks were tested on groups of children which were not strictly comparable.

Secondly, the selection of the children was left to the Head Teacher of the school and was made on the principle that both "controls" and "feeders" should be representative of the average children between 5 and 12 years of age: the actual method of selection being important I quote from Drs Leighton and McKinlay's * Report: "The teachers selected the two classes of pupils, those getting milk and those acting as "controls," in two different ways. In certain cases they selected them by ballot and in others on an alphabetical system." So far so good, but after invoking

* Department of Health for Scotland. *Milk Consumption and the Growth of Schoolchildren.* By Dr Gerald Leighton and Dr Peter L. McKinlay. (Edinburgh and London: H.M. Stationery Office, 1930.)

the goddess of chance they unfortunately wavered in their adherence to her for we read: "In any particular school where there was any group to which these methods had given an undue proportion of well fed or ill nourished children, others were substituted in order to obtain a more level selection." This is just the sort of afterthought that most of us have now and again and which is apt to spoil the best laid plans. In this case it was a fatal mistake, for in consequence the controls were, as pointed out in the Report*, definitely superior both in weight and height to the "feeders" by an amount equivalent to about 3 months' growth in weight and 4 months' growth in height.

Presumably this discrimination in height and weight was not made deliberately, but it would seem probable that the teachers, swayed by the very human feeling that the poorer children needed the milk more than the comparatively well to do, must have unconsciously made too large a substitution of the ill-nourished among the "feeders" and too few among the "controls" and that this unconscious selection affected, secondarily, both measurements.

Thirdly, it was clearly impossible to weigh such large numbers of children without impedimenta. They were weighed in their indoor clothes, with certain obvious precautions, and the difference in weight between their February garb and their somewhat lighter clothing in June is thus necessarily subtracted from their actual increase in weight between the beginning and end of the experiment. Had the selection of "controls" and "feeders" been a random one, this fact, as pointed out in the Report*, would have mattered little, both classes would have been affected equally, but since the selection was probably affected by poverty it is reasonable to suppose that the "feeders" would lose less weight from this case than the "controls." It is therefore not surprising to find that the gain in weight of "feeders" over "controls," which includes this constant error, was more marked, relatively to their growth rate, than was their gain in height, which was fortunately not similarly affected.

Fourthly, the "controls" from those schools which took raw milk were bulked with those from the schools which took pasteurised milk.

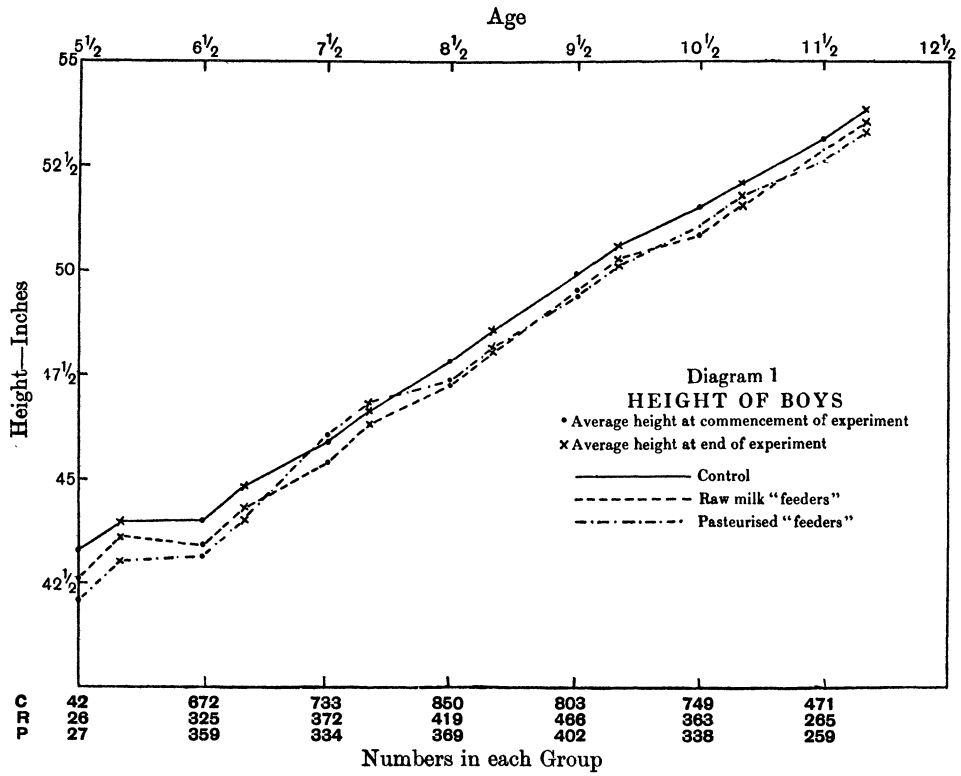
Now with only 67 schools, at best 33 against 34, in a district so heterogeneous both racially and socially, it is quite possible that there was a difference between the averages of the pupils at 33 schools and those of the pupils at another 34 schools both in the original measurements and in the rate of growth during the experiment.

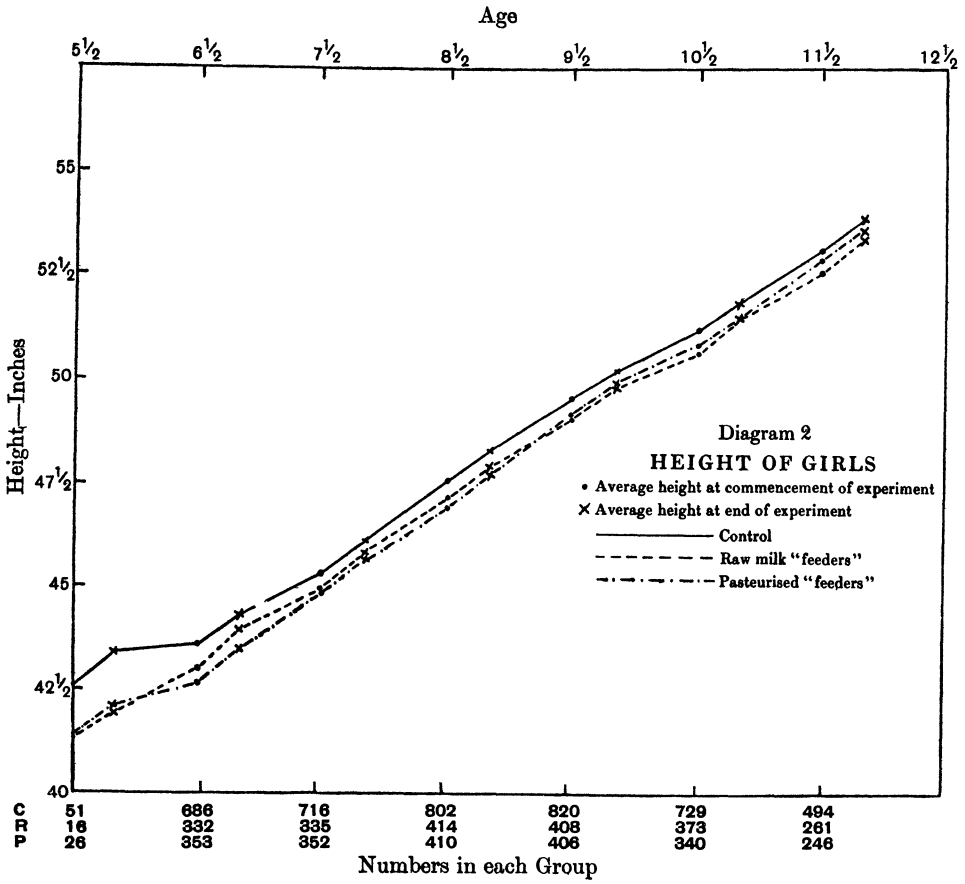
In that case the average "control" could not be used appropriately to compare with either the "raw" group or the "pasteurised" group.

This possibility is enhanced by the aforementioned selection of "controls" which can hardly have been carried out in a uniform manner in different schools.

Fortunately it would still be possible to correct this, for the figures for the different schools must still be available in the archives.

* See footnote on p. 398.





Diagrams 1 and 2 give the average heights of "controls," raw milk "feeders" and pasteurised milk "feeders" for boys and girls respectively. The heights at the beginning of the experiments are set out against a uniform age scale centring each group at the half year above the whole number. This is doubtless accurate enough except for the first group aged "5 and less than 6," which was very much smaller in numbers than the other groups, either because only the older (or larger) children are sent to school between 5 and 6 or because the teachers did not think that the smaller children would be able to play their part. For this reason they should probably be centred more to the right compared to the others. A similar argument might lead us to centre the "11 and over" group a little more to the left.

The average heights at the end of the experiment are of course set out four months to the right of those at the beginning and it will be noticed that except for the first group, which is clearly out of place, not any of the points diverge very much from their appropriate line of growth whether "controls," "raws" or "pasteuriseds."

The case is very different in Diagrams 3 and 4 which show the corresponding average weights. Here there is, after the first two ages, a very decided dip, especially in the later ages. The weights at the end of the experiment are too low. This might be accounted for by a tendency in older children to grow normally in height and subnormally in weight during the spring, but I think it much more likely that older children wear about 1 lb. more clothes in February than they do in June, while in the case of younger children a more limited wardrobe permits of fewer discards.

The authors have tried to show that the selection of the "controls" has not affected the validity of the comparison, by computing the correlation coefficients between the original heights (and weights) and the growth during the experiment for each of the 42 age groups into which the measurements were divided. These they find to be quite small even though they are here and there significant, and they argue that the additional height and weight of the "controls" was without effect on the comparison of subsequent growth.

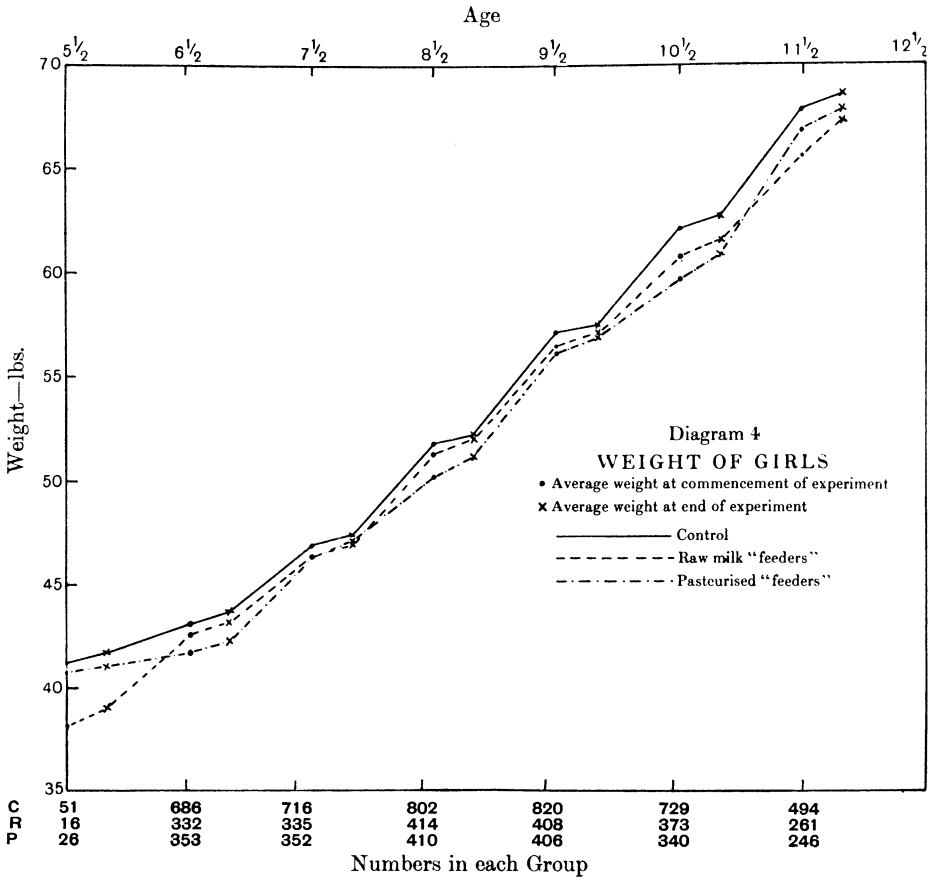
Now this might have been a perfectly good argument had the height and weight been selected directly, but if, as I have indicated was very likely the case, the selection was made according to some unconscious scale of well being, then it is surely natural to suppose that the relatively ill nourished "feeders" would benefit more than their more fortunate school mates, the "controls," would have done by the extra $\frac{3}{4}$ pint of milk per day.

That being so how are we to regard the conclusions of the Report*:

(1) "The influence of the addition of milk to the diet of school children is reflected in a definite increase in the rate of growth both in height and weight."

This conclusion was probably true; the average increase for boys' and girls' heights was 8 per cent. and 10 per cent. over "controls" and for boys' and girls' weights was 30 per cent. and 45 per cent., respectively, and though, as pointed out,

* See footnote on p. 398.



the figures for weights were wholly unreliable it is likely enough that a substantial part of the difference in height and a small part of that in weight were really due to the good effect of the milk. The conclusion is, however, shifted from the sure ground of scientific inference to the less satisfactory foundation of mere authority and guesswork by the fact that the "controls" and "feeders" were not randomly selected.

(2) "There is no obvious or constant difference in this respect between boys and girls and there is little evidence of definite relation between the age of the children and the amount of improvement. The results do not support the belief that the younger derived more benefit than the older children. As manifested merely by growth in weight and height the increase found in younger children through the addition of milk to the usual diet is certainly not greater than, and is probably not even as great as, that found in older children."

Now from the authors' point of view, believing in the validity of their comparisons in weight, this is much understating the case, as the following table derived from Capt. Bartlett's condensed tables shows:

Age in years	Gain in weight in ozs. by Feeders over Controls		Gain in height in inches by Feeders over Controls		As °/₀ of control			
					Weight		Height	
	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
5, 6 and 7	1·13 ± ·73	1·24 ± ·72	·083 ± ·011	·059 ± ·011	9	13	11	8
8 and 9	3·15 ± ·68	4·47 ± ·67	·071 ± ·011	·098 ± ·010	30	51	10	14
10 and 11	5·21 ± ·85	7·88 ± ·79	·037 ± ·012	·055 ± ·012	78	73	5	8

Note that the P.E.'s are calculated from Capt. Bartlett's tables and are subject, as his are, to his having interpreted the methods of the original Report correctly.

From this they might have concluded:

(*a*) That in the matter of weight older children, both boys and girls, derived more benefit than younger, while

(*b*) In height the younger boys did better than the older, though the difference is not quite significant, but that there was no regular tendency in the matter of girls' height.

In the light of previous criticism, however, we must be content to say that apparently the differential shedding of clothes between the "feeders" and the more fortunate "controls" is more marked with older children (and possibly with girls than with boys), and that there is some probability that younger boys gain in height more than older.

Finally, conclusion (3) runs: "In so far as the conditions of this investigation are concerned the effects of raw and pasteurised milk on growth in weight and height are, so far as we can judge, equal."

This conclusion has been challenged by Capt. Bartlett*, and by Dr Fisher and Capt. Bartlett†, who conclude that there is definite evidence of the superiority of raw over pasteurised milk in both height and weight.

Even they however point out that the raw and pasteurised milk were not supplied to the same schools, and their conclusion amounts to saying: "If the groups of children taking raw and pasteurised milk respectively were random samples from the same population, the observed differences would be decisively in favour of the raw milk."

Unfortunately they were not random samples from the same population: they were selected samples from populations which may have been different, and moreover the "controls" with which they were compared were not appropriate to either group; and so—again it is a matter of guess and authority—I would be very chary of drawing any conclusion from these small biased differences.

That is not to say that there is no difference between the effect of raw and pasteurised milk—personally I believe that there is and that it is in favour of raw milk—but that this experiment, in spite of all the good work which was put into it, just lacked the essential condition of randomness which would have enabled us to prove the fact.

This note would be incomplete without some constructive proposals in case it should be considered necessary to do further work upon the subject, and accordingly I suggest the following:

(1) If it should be proposed to repeat the experiment on the same spectacular scale,

(*a*) The "controls" and "feeders" should be chosen by the teachers in pairs of the same age group and sex, and as similar in height, weight and especially physical condition (i.e. well or ill nourished) as possible, and divided into "controls" and "feeders" by tossing a coin for each pair. Then each pair should be considered to be a unit and the gain in weight and height by the "feeder" over his own "control" should also be considered as a unit for the purpose of determining the error of the gain in weight or height.

In this way the error will almost certainly be smaller, perhaps very much smaller, than if calculated from the means of "feeders" and "controls."

If in addition the social status of each pair be noted (well to do, medium, poorly nourished or some such scale) further useful information will be available for comparing pasteurised and raw "feeders."

If this is found to be too difficult a perfectly good comparison can be made by adhering to the original plan of the 1930 experiment and drawing lots to decide which should be "controls" and which "feeders" (this is better than an alphabetical arrangement), but the error of the comparison is likely to be larger than in the plan outlined above.

* "Nutritional Value of Raw and Pasteurised Milk," by Stephen Bartlett, M.C., B.Sc. (*Journal of the Ministry of Agriculture*, April, 1931).

† *Nature*, April 18th, 1931, p. 591, "Pasteurised and Raw Milk."

(*b*) If it is at all possible each school should supply an equal number of raw and pasteurised "feeders," again by selection of similar children followed by coin tossing, but I fear that this is a counsel of perfection.

(*c*) Some effort should be made to estimate the weight of clothes worn by the children at the beginning and end of the experiment: possibly the time of year could be chosen so that there would be little change in this respect.

(2) If it be agreed that milk is an advantageous addition to children's diet—and I doubt whether any one will combat that view—and that the difference between raw and pasteurised milk is the matter to be investigated, it would be possible to obtain much greater certainty at an expenditure of perhaps 1—2 per cent. of the money* and less than 5 per cent. of the trouble.

For among 20,000 children there will be numerous pairs of twins; exactly how many it is not easy to say owing to the differential death rate, but, since there is about one pair of twins in 90 births, one might hope to get at least 160 pairs in 20,000 children. But as a matter of fact the 20,000 children were not all the Lanarkshire schools population, and I feel pretty certain that some 200—300 pairs of twins would be available for the purpose of the experiment.

Of 200 pairs some 50 would be "identicals" and of course of the same sex, while half the remainder would be non-identical twins of the same sex.

Now identical twins are probably better experimental material than is available for feeding experiments carried out on any other mammals, and the error of the comparison between them may be relied upon to be so small that 50 pairs of these would give more reliable results than the 20,000 with which we have been dealing.

The proposal is then to experiment on all pairs of twins of the same sex available, noting whether each pair is so similar that they are probably "identicals" or whether they are dissimilar.

"Feed" one of each pair on raw and the other on pasteurised milk, deciding in each case which is to take raw milk by the toss of a coin.

Take weekly measurements and weigh without clothes.

Some way of distinguishing the children from each other is necessary or the mischievous ones will play tricks. The obvious method is to take finger-prints, but as this is identified with crime in some people's minds, it may be necessary to make a different indelible mark on a fingernail of each, which will grow off after the experiment is over.

With such comparatively small numbers further information about the dietetic habits and social position of the children could be collected and would doubtless prove invaluable.

The comparative variation in the effect in "identical" twins and in "unlike" twins should furnish useful information on the relative importance of "Nature and Nurture."

* This is a serious consideration: the Lanarkshire experiment cost about £7500.

To sum up: The Lanarkshire experiment devised to find out the value of giving a regular supply of milk to children, though planned on the grand scale, organised in a thoroughly business-like manner and carried through with the devoted assistance of a large team of teachers, nurses and doctors, failed to produce a valid estimate of the advantage of giving milk to children and of the difference between raw and pasteurised milk.

This was due to an attempt to improve on a random selection of the controls which in fact selected as controls children who were on the average taller and heavier than those who were given milk.

The hypothesis is advanced that this was due not to a selection of the shorter, lighter children as such to take the milk, but to an unconscious bias leading the teachers to pick out for this purpose the needier children whom the milk would be most likely to benefit.

This hypothesis is supported by the fact that while the advantage derived from the milk was only 8—10 per cent. of the gain in height, without much variation for age, it was 30—45 per cent. of the gain in weight, varying from 9—13 per cent. in the younger children (who do not seem to have shed much clothing in the summer) up to 73—78 per cent. in the older children—who obviously did.

Suggestions are made for the arrangement:

(1) Of a similar large scale experiment on random lines, and

(2) Of a much smaller and cheaper experiment carried out on pairs of twins of like sex.

The second is likely to provide a much more accurate determination of the point at issue, owing to the possibility of balancing both nature and nurture in the material of the experiment.