THE CONDUCT OF CLINICAL TRIALS OF SUBSTANCES
PROPOSED FOR THE NUTRITION OF
INFANTS AND CHILDREN

Including a Critique of the Nitrogen Balance Technique
for Measuring Protein Anabolism

By Charles U. Lowe, M.D., and Vivian Pessin, M.A.
Department of Pediatrics, School of Medicine, University of Buffalo

Animal experiments may offer evidence of the value of a substance for human nutrition, but only well-conducted clinical trials can result in a verdict. Unless substances are tested on human subjects, judgment should be withheld. Though the design of experiments varies, certain general principles pertain to all experimentation; in addition, special considerations apply in the field of nutrition. The outline which follows presents some of these principles and considerations and should be useful to those conducting clinical trials.

A statement of specific objectives should be prepared as the first step. In order to estimate the necessary sample size, estimates of expected results must be obtained, or deduced, either from the literature or from a pilot experiment. All possible sources of bias, such as age, sex, race, etc. should be listed in advance and considered in planning the study. Decisions must also be made concerning the following: 1) population for which the results will be relevant; 2) controls; 3) criteria for judging results; 4) sample size and duration of study; 5) form of reporting results. Whenever possible the people who give treatments and make measurements should be unable to distinguish between the study and the control subjects.

1. POPULATION

The type of population for which the conclusion is intended will determine the choice of subjects. For example the value of a foodstuff for all infants should be studied in subjects from a heterogeneous population. Alternatively, a dietary supplement for treating deficiency must be assayed in deficient subjects. Such a supplement would probably produce no significant changes in normal subjects but even if it did, the results would not necessarily apply to treatment of deficiency.

2. CONTROLS

Controls, necessary in most clinical tests, should come from the same population as the treated group. When treated individuals are used as their own controls, the trial periods should, when possible, be both preceded and followed by a control period.

3. CRITERIA

The nature and objective of the clinical trial should determine the choice of criteria and measurements. These may be: a) state of health; b) growth of specialized structures; c) maturation; d) direct determination of growth; e) indirect determination of growth.

a. Health

The aim of good nutrition is, of course, the achievement and maintenance of optimal health. Unfortunately, there are neither ideal standards of, nor simple methods for, measuring health. Indirect indices, such as total hemoglobin, total serum protein or antibody production may be useful. The incidence of morbidity in study groups as evidence of the absence of health is also informative.1

ADDRESS: (C.U.L.) Children’s Hospital, 219 Bryant Street, Buffalo 22, New York.

694
b. Growth of Specialized Structures

Certain tissues have growth requirements which are qualitatively different from those of protoplasm. For example, more sulphur-containing amino acid is needed for hair growth than for protoplastic anabolism. It is possible that on a diet marginal with respect to such a constituent, the growth of specialized structures will suffer before over-all growth is affected. Nail growth has been proposed as a sensitive index of this sort.²

c. Maturation

Norms of rates of epiphyseal maturation³ are available and these, as well as length of long bones⁴ and dentition⁵ have been used to evaluate nutrition. The correlation of these indices with height and weight may be poor. For example, epiphyseal calcification was found to progress more slowly than weight in malnourished children treated with protein supplements.⁶ Furthermore, in patients with anemia due to iron deficiency, height and weight may be reasonably normal while epiphyseal maturation may be retarded.

d. Direct Determination of Growth

As growth is the result of related changes in weight and height, these may well be studied together. Ideal patterns for growth of healthy children are not known, though many anthropometric studies provide data of growth norms for selected populations.⁷-⁸ It is moot, however, whether growth at a rate greater than the norms in published tables is beneficial. Increased consumption of nutritious diets may accelerate weight gain of normal infants; such increase may be temporary, because these infants often, at some later time, stop growing while chronologic age advances to morphologic age.⁹ If, as implied by Wetzel’s data,¹⁰ the eventual maximum size and habitus of normal children are independent of temporary changes in rate of growth, a diet which produces transient acceleration of growth will not affect ultimate stature.

Because there are no universally accepted standards for ideal growth, each study should provide appropriate growth standards from controls and avoid comparing growth of the study group to arbitrary growth standards. For example, growth of a control group of infants receiving human milk plus iron and vitamin supplements is a useful standard in the first 6 months of life. However, measurement of total weight gain does not distinguish between protein synthesis and accumulation of fat or water.

e. Indirect Determination of Growth

The use of measurements which change more rapidly than height, weight and indices of maturation may shorten the time needed to evaluate the effect of a substance or diet. One such measurement is N retention as determined by balance techniques, as all growth, though not all weight gain, can ultimately be related to protein synthesis. However, the uncritical use of this technique is of less value than simple measurements of weight gain.

Protein anabolism implies a positive balance of N, a gain in weight, and the retention of P, S and K in the same proportions as they exist in protoplasm.¹¹ Balance data on infants obtained before 1925 fail to show consistent quantitative relationship between the retention of N and other constituents of protoplasm.¹²,¹³ This inconsistency may have resulted from poor diets, inadequate intake of vitamins A, D and C, insufficient collection periods, or faulty chemical methods. Some N balance studies suggest that diets high in minerals and nitrogen yield high retentions and apparent “supermineralization.”¹⁴,¹⁵ These studies, however, neglect to consider salt loss through the skin;¹⁶-¹⁸ correction for this loss might have appreciably lowered the apparent retentions. A more recent study has indicated that the high retention of salt on mineral-rich diets is temporary.¹⁹ Nevertheless, it is also possible that the composition of protoplasm may be, within limits, affected by diet.

The normal fluctuation in retention is
However, either long-term observation of a single child or shorter observations of many children often show a correspondence between observed weight gain and theoretical weight gain calculated from retention of \( N, S, K \) and \( P \) (corrected for \( \text{Ca} \) retention) (Table I). Although the conversion factors used in this table are inexact and make no correction for weight gain due to fat accumulation or bone growth, they are useful in making a first approximation of the relationship between these variables.

In spite of the difficulties of obtaining and interpreting nitrogen balance data from infants and children, it is possible to suggest an experimental plan which is likely to yield useful information. A fore-period is necessary to insure that the subject is in a steady state. The length of this period will depend on the metabolite being studied and the size of the difference between the diet under investigation and the preceding diet. Adjustment to an abrupt change in intake requires time; change from very low to very high intake may lead to large temporary retention and vice versa.\(^{11,20}\) Three-day pools of feces and urine are convenient both for chemical analysis and for evaluation of data. The necessary number of collection periods should be determined by the nature of the experiment; a more specific and detailed treatment of this question is given below (cf. Sample Size and Duration of Study). The use of published tables on composition of foods is inadvisable.\(^{31}\) Diet, feces and urine must be analyzed chemically and feces should be collected using some type of marker to delineate periods. The publication of basic data, such as variations in replicate determinations and recovery experiments, will demonstrate the degree of precision of laboratory procedures and contribute considerably to the acceptability of data.

At least one substance other than nitrogen, from which a theoretical nitrogen retention can be calculated, should be measured. In theory, sulphur should be good for this purpose because it is retained almost solely as a constituent of body protein and is not lost from the skin in any great quantity; neither phosphorus nor potassium offers both of these advantages. Acceptable

### TABLE I

<table>
<thead>
<tr>
<th>No. Pts.</th>
<th>Age</th>
<th>Diet</th>
<th>Days</th>
<th>Weight Gain (gm/day)</th>
<th>Theoretical, Based on Retention of:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Total Obs.</td>
<td>Per Period</td>
<td>( N )</td>
</tr>
<tr>
<td>1</td>
<td>1-3 mo</td>
<td>1</td>
<td>90</td>
<td>6</td>
<td>25.7</td>
</tr>
<tr>
<td>1</td>
<td>1-3 mo</td>
<td>2</td>
<td>90</td>
<td>6</td>
<td>28.3</td>
</tr>
<tr>
<td>5</td>
<td>4-7 mo</td>
<td>3</td>
<td>77</td>
<td>7</td>
<td>33.8</td>
</tr>
<tr>
<td>4</td>
<td>8-34 wk</td>
<td>1</td>
<td>72</td>
<td>3</td>
<td>22.7</td>
</tr>
<tr>
<td>9</td>
<td>4y-8 yr</td>
<td>4</td>
<td>2010</td>
<td>5</td>
<td>11.0</td>
</tr>
<tr>
<td></td>
<td>Unweighted average:</td>
<td></td>
<td></td>
<td></td>
<td>23.8</td>
</tr>
</tbody>
</table>

* Averages of pooled data. Calculations performed assuming: 1) no change in glycogen content of body; 2) all weight gain due to increase in protoplasm. Calculations of theoretical weight gain made from measured retention by the following conversions:\(^{11}\)

\[
\text{K mEq} \quad \text{Ca gm} \\
N \times 32; \quad \text{——} \times 32; \quad P \text{ gm} = \quad \text{——} \times 32; \quad S \text{ gm} \times 14.5 \times 32. \\
2.17 \quad \frac{2}{2.29}
\]

† Diets: 1. cow's milk modifications; 2. human milk; 3. evaporated milk modification; 4. normal diet for age.

** Corrected for loss of \( K \) through skin.\(^{31}\)

‡ 10-day fore-periods.
nitrogen balance studies should show a reasonable quantitative relationship between nitrogen balance and weight change; furthermore, the theoretical nitrogen balance based on K, S or P should show reasonable quantitative relationship to the observed nitrogen balance. When these two conditions are met, it is likely that values for nitrogen retention correspond to synthesis of body protein and are due neither to a systematic analytic error nor to the accumulation of nonprotein nitrogen.

Diets designated to be deficient in one known nutrient have been used to evaluate the results of additions of this nutrient to the diet. Such experiments sometimes lead to anomalous or inconclusive results because many such diets are inadvertently deficient in other growth essentials. This is more likely to be true of diets intended to be deficient in one of the B vitamins or in protein. A related problem has appeared in the experimental determination of amino acid requirements for man. The minimal requirement of an essential amino acid has been shown to vary with the total amino acid content of the diet. Furthermore, effects on N balance of a group of amino acids depend on whether they are administered simultaneously or sequentially.

To summarize, the direct evaluation of health is difficult; however the absence of disease is useful and often obvious. There is too little information on the effect of diet on growth of specialized structures and on indices of maturation to judge the extent of correlation between these and other physical measurements, particularly in malnourished children. Of the criteria used to evaluate diets, the simplest and least liable to error is rate of weight gain. Unfortunately, the lack of sophistication of this method sometimes detracts from its appeal. Two real disadvantages are 1) the absence of a standard for ideal weight gain and 2) the difficulty of distinguishing between weight gain due to accumulation of fat or water and weight gain resulting from synthesis of protoplasm. Metabolic balance data may not have the second disadvantage but such data are valuable and reliable only if studies are carefully designed and results critically appraised.

4. SAMPLE SIZE AND DURATION

The purpose of sampling is to arrive at measurements which describe the population. Many sampling designs have been devised to fit different kinds of problems, but all of them are intended to accomplish the same objectives: 1) to obtain unbiased estimates of population values, and 2) to have estimates sufficiently precise for the purpose of the experiment.

Provision must be made for the first objective in the design of the experiment. Known possible sources of bias, such as age, weight, etc., should be considered and controlled by such devices as randomness, studying each group separately, and balancing the composition of the test and control groups with respect to each possible source of bias. Bias of unknown origin can be avoided by random selection of subjects.

The second objective may be achieved by a suitable choice of the size of the sample. Estimates from small samples are less precise than estimates from large samples. Modern statistical theory offers a method of determining a sample size in advance, so that the sample measurements can be expected to have a known precision. The procedures used take into account all causes of variation (except bias) that affect the test and control groups.

The omission of an advance decision on the necessary size of the sample often leads to inconclusive results. For example, it may be desired to determine whether or not a test procedure will result in an increase in weight. After the experiment is performed it is found that the test mean is larger than the control mean, but that the difference is not significant statistically. This result indicates one of two things: either there is no, or a relatively small, mean difference; or else a true and sizeable difference exists but the sample was too small to demonstrate this clearly. In the absence of other evidence, there is no way to decide which of
the two possible conclusions is the correct one.

For reasonable assurance that a conclusion will be reached at the end of an experiment, the following procedure is recommended. First the investigator should find (from the literature or from a pilot study) estimates of variation in the control and test groups, and also an estimate of the expected difference between these groups at the end of the experiment. Then he should decide on the level of significance he wants. From these it is possible to estimate the minimum number of observations necessary to give the expected difference the desired precision.

As the design of an experiment should be determined by its purpose and by the types of measurement used, there is no single design which will be best for all experiments. The example which follows was chosen to illustrate some of the principles mentioned above.

Suppose it is desired to determine whether a test diet will result in a higher N retention than a control diet. The sample of infants is to be divided at random into two equal groups, test and control. How large a sample, that is, how many children and how many balance periods will be required for a 10 or 20% difference in retention to be statistically significant?

Suitable estimates of variation were obtained by using data from Jeans (Table II).

The coefficient of variation ($V = \text{S.E.} \times 100/\text{mean}$) measures the magnitude of the scatter compared to the mean, and was found to be the same for the older groups. Therefore, for the purposes of this problem, the last two groups could be combined.

To compute the minimum number of observations necessary, the following procedure was used:

The basic formula is:

$$\text{Difference between means} \quad \frac{\text{S.E. of diff. between means}}{t} \geq \frac{\text{Difference of means}}{\text{n}^{\ast}}$$

For the 5- to 15-week-old group,

- Difference of means = 10% of mean = .101
- S.E. of difference of means = \( \sqrt{\frac{2}{n}} \) \( .44 (1.414) \) \( .62 \)

where \( n \) is the number in the test group and also in the control group.

\( t \) (for samples over 30*) = 1.96, for a 5% level of significance.

Then, substituting in the basic formula,

$$\frac{.101}{.62} \geq 1.96$$

or $\sqrt{n} \geq \frac{1.96 (.62)}{.101} = 12.03$

and $n \geq 145$

Thus for a 10% difference in means at age 5 to 15 weeks to be significantly different at the 5% level of significance, at least 145 balance periods will be needed both in the test and in the control groups. These can be obtained, in theory, by taking 145 periods for each of two children, or 15 periods per child with 20 children or any other combination of number of children and number of balance periods which will

* The example given is not to be construed as a model or pattern for clinical trials, i.e., the number of subjects and periods of observation are only suitable for the conditions specified and cannot be taken as generally suitable or necessary for any clinical trial.

**TABLE II**

**ESTIMATES OF MEANS, STANDARD ERRORS, AND COEFFICIENTS OF VARIATION OF N BALANCE, BY AGE**

(for illustrative example discussed in text)

| Age (wk) | No. of 3-day Periods | N Balance (gm/day) | Mean | S.E. | V
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5-15</td>
<td>31</td>
<td>1.01</td>
<td>.44</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>16-25</td>
<td>45</td>
<td>.97</td>
<td>.34</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>26-45</td>
<td>28</td>
<td>1.28</td>
<td>.42</td>
<td>34</td>
<td></td>
</tr>
</tbody>
</table>

* Based on data from Jeans'5
† Coefficient of variation.

* For t-table, see R. A. Fisher: Statistical Methods for Research Workers. For samples under 30, n can be determined by trial and error.
give a minimum of 145 periods for each group. For simplicity, it was assumed in the example that the variations between children, between balance periods, and in the test and control groups are all equal. If they are substantially different, and if estimates of all are available, the calculations should be performed using all estimates of variation.

Table III lists the minimum number of balance periods theoretically necessary in the example chosen for statistical significance of the difference of two means, in different age groups for various combinations of numbers of children, expected magnitude of difference, and level of significance. In practice, some of these plans may not work out. For example, it is obviously impossible to plan an experiment using only one test and one control child from 5 to 15 weeks of age because the children would be out of the age group before the required 145 balance periods could be obtained. If, for any reason, all the combinations of children and balance periods are impracticable, consideration should be given to modifying the design of the experiment. A modification which would result in an increase in the expected change in the N balance would decrease the minimum number of children-periods required.

It should be emphasized that the numbers of observations obtained by this method are minimal, as no allowance is made for an observed difference in means being smaller, or the observed variations being larger, than the expected values.

In the example chosen above, the duration of the experiment was calculated to provide the minimum number of observations necessary to demonstrate statistical significance of expected differences. All of the variables dealt with here will be present in any nutritional experiment even if balance techniques are not used. In certain experiments, the period of observation should be much longer than the minimum necessary for statistical significance. For example, if it is desired to determine whether an observed difference will persist for a number of years, the observations must continue for that period of time.

### Table III

Example: Minimum Number of Children and 3-Day Balance Periods Required for the Specified Level of Statistical Significance of the Difference Between Two Means

(discussed in text)

<table>
<thead>
<tr>
<th>No. of Children (half test, half control)</th>
<th>10% Difference Expected in N Balance</th>
<th>20% Difference Expected in N Balance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5% Level of Signif.</td>
<td>2% Level of Signif.</td>
</tr>
<tr>
<td>Age 5–15 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>145</td>
<td>203</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>41</td>
</tr>
<tr>
<td>20</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>40</td>
<td>8</td>
<td>11</td>
</tr>
<tr>
<td>60</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>Age 16–45 weeks</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>92</td>
<td>129</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>26</td>
</tr>
<tr>
<td>20</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>40</td>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>60</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

* Based on data from Jeans (Table II).
5. FORM OF REPORTING RESULTS

The sample and the method by which it was chosen should be specified. The description of diets used, laboratory methods, etc., should be in sufficient detail so that the experiment could be duplicated without additional information. The original data should be included; these will not only support the conclusions but will also be valuable to other investigators. Statistical analysis appropriate to the objectives and design of the experiment is necessary, including such summary observations as the mean and standard deviation, statistical tests used and their results and the level of significance. Statements made should be supported by the text or by references, or characterized as opinion. Conclusions should be accompanied by a statement of the population to which they apply, limiting conditions, and degree of certainty.

REFERENCES

THE CONDUCT OF CLINICAL TRIALS OF SUBSTANCES PROPOSED FOR
THE NUTRITION OF INFANTS AND CHILDREN: Including a Critique of the
Nitrogen Balance Technique for Measuring Protein Anabolism
Charles U. Lowe and Vivian Pessin
Pediatrics 1957;19:694

Updated Information & Services
including high resolution figures, can be found at:
http://pediatrics.aappublications.org/content/19/4/694

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
https://shop.aap.org/licensing-permissions/

Reprints
Information about ordering reprints can be found online:
http://classic.pediatrics.aappublications.org/content/reprints
THE CONDUCT OF CLINICAL TRIALS OF SUBSTANCES PROPOSED FOR
THE NUTRITION OF INFANTS AND CHILDREN: Including a Critique of the
Nitrogen Balance Technique for Measuring Protein Anabolism
Charles U. Lowe and Vivian Pessin
Pediatrics 1957;19:694

The online version of this article, along with updated information and services, is located on
the World Wide Web at:
http://pediatrics.aappublications.org/content/19/4/694