A method for the rapid assessment of sample size in dietary studies[1,2]

John C Hall MB, BS, FRACS

ABSTRACT Critical readers should be suspicious about the inability of a dietary study to discriminate between the energy intakes of two groups when small sample sizes have been used. The possibility of a false-negative (type II error) should be considered. This problem could be avoided if investigators used adequate sample sizes. A review of 26 dietary studies published in the *American Journal of Clinical Nutrition* between 1979 and 1981 revealed that the median "SD of energy intakes" was 525 kcal/day. This figure was used to illustrate a simple method for estimating appropriate sample sizes assuming type I and type II error probabilities of 0.05. Prospective use of this method should increase the reproducibility of conclusions drawn from dietary studies. *Am J Clin Nutr* 1983;37:473–477.

KEY WORDS Diet, energy intake, sample size, type II errors, statistical power

Introduction

Too much confidence is placed on negative results derived from small sample sizes (1). Perhaps the most common design error, wrote Altman (2) when commenting on clinical trials, is to have too small a sample to get reliable and/or useful results. A possible serious consequence of having too small a sample (and thus a lack of statistical "power") is the inability to detect an important effect.

The purpose of this paper is to present a method for the estimation of sample sizes in dietary studies.

The background

Tests of statistical significance do not take into account the strength of the relationship between the variables under study: a difference in response rates of 50% may fail to be significantly different if the numbers of subjects studied are too small, whereas a difference of as little as 1% may be significant if large enough numbers are studied. It is this fact that has led Sackett (3) to describe a "wrong sample size bias," ie, samples that are too small can prove nothing, samples that are too large can prove anything.

At present there is a widespread bias toward the use of small sample sizes. Der Simonian et al (4) considered 11 important aspects of design and analysis in 67 clinical trials published in major medical journals. The statistical power of the study to detect treatment effects was discussed in only 12% of the articles. It is evident that few investigators give adequate attention to sample size when planning studies. As a result the possibility of false-negative (type II errors) is considerable.

The concept of type II errors can be explained as follows. The first step in testing the significance of an observed difference between two groups is to formulate a null hypothesis, ie, that there is, in reality, no differences between the groups under study. If the null hypothesis can be rejected, then the experimenter will conclude that a real difference is present. Rejection of the null hypothesis when, in fact, it is true is a type I error (**Table 1**): the probability of making this type of error is α and is usually set at less than one in 20 ($p < 0.05$). In contrast, rejection of the null hypothesis when, in fact, it is false is a type II error (Table 1): the probability of making this type of error is β.

Intuition dictates that the larger the sample size, the easier it is to be certain of displaying any particular degree of difference between

[1] From the Department of Surgery, The Flinders University of South Australia, Bedford Park, Australia.

[2] Reprints will not be available.

Received June 7, 1982.
Accepted for publication September 21, 1982.

TABLE 1
The possible outcomes when two groups are compared

	Probability of detecting a difference	Probability of not detecting a difference
No difference between the groups (null hypothesis true)	α (Type I error) False-positive	$1 - \alpha$ (Confidence) True-negative
A difference between the groups (null hypothesis false)	$1 - \beta$ (Power) True-positive	β (Type II error) False-negative

the groups under study. The smaller the degree of difference considered as useful, the larger the numbers that will be necessary to display this degree of difference. It should follow logically that authors who claim that a statistical difference is present should declare the false-positive error rate (α), and that authors who claim that no statistical difference is present should declare both the difference that they consider to be of clinical interest, and the false-negative error rate (β). Unfortunately few authors document the strength of their claims when describing negative results. They fail to appreciate that the terms "no significant difference" and "an insignificant difference" are not equivalent.

In a study by Freiman et al (5) of 71 negative randomized control trials, 67 of the trials had a greater than 10% risk of missing a true 25% therapeutic improvement. But the problem is not just restricted to clinical trials, it applies to all analytical studies. One study found that 60% of negative analytical studies had a greater than 50% chance of missing a true 10% difference between the groups under study (6).

Such problems could be overcome if investigators designed studies that included adequate sample sizes. Before presenting a method for the determination of sample sizes applicable to dietary studies I will briefly review some published dietary information. The only aim of this exercise is to obtain some estimates of energy intake that will help to simplify the method that is to be presented.

The review

Dietary studies published in the *American Journal of Clinical Nutrition* between 1979 and 1982 were reviewed. Twenty-six publications (7–32) contained subjects that were fed ad libitum, and had mean energy intakes between 1000 and 4000 kcal/day. The dietary methods used included food-intake records, recall histories, and food frequencies. The period of time used to collect this information varied from 1 to 7 days. Only the first set of data encountered in each study was selected for use.

The information derived from these studies could be summarized as a median "mean energy intake" of 1929 kcal/day (Interquartile range 1611 to 2250 kcal/day); a median "SD of energy intake" of 525 kcal/day (Interquartile range 389 to 646 kcal/day); and a median study number of 46 subjects (Interquartile range 20 to 95). A typical study could therefore be summarized as 1929 ± 525 kcal/day based on 46 subjects.

For the purpose of the estimates to be made in the following section it is important to note that there was little correlation between the mean energy intakes and the SD of energy intakes ($r = 0.15$), or between the SD of energy intakes and the numbers of subjects studied ($r = 0.12$) using the Spearman rank correlation coefficient (33). The influence of other factors on the SD of energy intake will also be discussed in the next section.

The estimate

The estimated sample size for a study comparing two independent means may be derived from the following formula (34):

$$n = 2 \cdot \left[\frac{(Z_\alpha - Z_\beta)\sigma}{\delta} \right]^2$$

where, n = the sample size in each group,
Z_α = the upper α percent point of a normal distribution (α is the probability of a false-positive error),
Z_β = the lower β percent point of a normal distribution (β is the probability of a false-negative error),
σ = the SD of the variable under study; based on the assumption that it is approximately the same for both groups,
δ = the difference in population

means that is thought to be of interest.

This formula can be simplified if it is assumed that a β of 0.05 and an α of 0.05 are of interest, ie, a one in 20 chance of either a false positive or a false-negative result is acceptable. For a one-tailed test Z_α is then fixed at 1.65 (the point cutting off 5% in the upper tail of the normal distribution), and Z_β is fixed at -1.65 (the point cutting off 5% in the lower tail of the normal distribution). The simplified formula is:

$$n = 2 \cdot \left[\frac{3.3 \cdot \sigma}{\delta}\right]^2$$

This formula can now be solved if the investigators can declare the difference in means between the groups that is of interest to them (δ), and estimate a realistic value for the standard deviation (σ). There are several ways for estimating the SD (35). *1*) It may be based on prior experience, eg, the value currently obtained in similar published studies (hence our estimate of 525 kcal/day which appears to be reasonable for studies of energy intake with mean values between 1000 and 4000 kcal/day), or a value obtained in the past by the investigators when studying a similar population; *2*) a pilot study may be performed; or *3*) a preliminary sample from the study proper may be used, eg, the first 50 patients entered into the study may be used for this purpose. If this strategy is used then any possible observer or interpreter bias should be avoided by having the calculations performed by someone external to the study.

It is acceptable to pool the results of both groups under study when making these calculations. Prior experience indicates that the SDs of energy intakes for different populations are usually about the same regardless of the amount of energy consumed within the range of 1000 to 4000 kcal/day. Furthermore it is desirable to adopt one of the methods that approximates the study conditions. This is because the magnitude of the SD may vary with the different techniques used to determine energy intake, and the period of time used to collect the data.

If, however, we assume that 525 kcal/day, as found from a review of the literature, is a reasonable estimate of the anticipated SD

TABLE 2
Estimates for the size of samples to be used in studies of energy intake (given the assumptions that $\alpha = 0.05$, $\beta = 0.05$, one-tailed test, and SD of both groups = 525 kcal/day)

Difference in the mean values of energy intake (kcal/day) between the groups that are of interest (δ)	Sample size for each group (n)	Total sample size (2n)
50	2401	4802
100	600	1200
150	267	534
200	150	300
250	96	192
300	67	134
350	49	98
400	38	76
450	30	60
500	24	48
600	17	34
700	12	24
800	9	18
900	7	14
1000	6	12

then the formula reduces to:

$$n = 2 \cdot \left[\frac{3.3 \cdot 525}{\delta}\right]$$

which approximates to:

$$n = \frac{6{,}000{,}000}{\delta^2}$$

It is now possible to estimate the sample size given any set difference in mean energy intake between the groups that might be of interest. This has been performed over a range of values in **Table 2.**

Some investigators may believe that fixing the error probabilities of α and β at 0.05 is too stringent. An alternative is to keep α at 0.05 and set β at 0.10 or 0.20. This course of action is to compromise between what is logistically obtainable and what is necessary to minimize the possibility of a false-negative error.

If despite these measures a negative result does occur then the power of this finding can be documented by calculating the probability of a type II error. Feinstein (36) and Fleiss (37) have described the methods for doing this.

Conclusion

Any investigator designing a study that compares the energy intakes of two groups

can plan ahead by estimating the sample size needed before the commencement of the study. This will ensure that the results are statistically reliable, and will diminish the possibility of a false-negative result. A method for the estimation of sample sizes in dietary studies has been presented herein. Although energy intake data have been used the same approach is applicable to studies that investigate other food components.

References

1. Altman DG. Statistics in medical research. III. How large a sample? Br Med J 1980;281:1336–8.
2. Altman DG. Statistics in medical journals. Statistics in medicine. 1982;1:59–71.
3. Sackett DL. Bias in analytical research. J Chron Dis 1979;32:51–63.
4. Der Simonian R, Charette J, McPeek B, Mosteller F. Reporting on methods in clinical trials. N Engl J Med 1982;306:1332–7.
5. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error, and sample size in the design and interpretation of the nonrandomized control trial. N Engl J Med 1978;299:690–4.
6. Hall JC. The other side of statistical significance: A review of type II errors in the Australian medical literature. Aust NZ J Med 1982;12:7–9.
7. Hepner GH, Fried R, St Jeor S, Fusetti L, Manin R. Hypocholesterolemic effect of yogurt and milk. Am J Clin Nutr 1979;32:19–24.
8. Mora JO, de Paredes B, Wagner M, et al. Nutritional supplementation and the outcome of pregnancy. I. Birth weight. Am J Clin Nutr 1979;32:455–62.
9. Vuori L, Christiansen N, Clement J, Mora JO, Wagner M, Herrera MG. Nutritional supplementation and the outcome of pregnancy. II. Visual habituation at 15 days. Am J Clin Nutr 1979;32:463–9.
10. Matkovic V, Kostial K, Simonovic I. Buzina R, Brodare A, Nordin BEC. Bone status and fracture rates in two regions of Yugoslavia. Am J Clin Nutr 1979;32:540–9.
11. Edozien JC, Switzer BR, Bryan RB. Medical evaluation of the special supplemental food program for women, infants, and children. Am J Clin Nutr 1979;32:677–92.
12. Jones RJ, Turner D, Ginther J, Brandt B, Slowie L, Lauger G. A randomized study of instructional variations in nutrition counseling and their efficacy in the treatment of hyperlipidemia. Am J Clin Nutr 1979;32:884–904.
13. Cerqueira MT, Fry M McM, Connor WE. The food and nutrient intakes of the Tarahumara Indians of Mexico. Am J Clin Nutr 1979;32:905–15.
14. Flynn MA, Nolph GB, Flynn TC, Kahrs R, Krause G. Effect of dietary egg on human serum cholesterol and triglycerides. Am J Clin Nutr 1979;32:1051–7.
15. Smith DA, Gee MI. A dietary survey to determine the relationship between diet and cholelithiasis. Am J Clin Nutr 1979;32:1519–26.
16. Vir SC, Love AHG. Nutritional status of institutionalized and noninstitutionalized aged in Belfast, Northern Ireland. Am J Clin Nutr 1979;32:1934–47.
17. Bronsgeest-Schoute DC, Hermus RJJ, Dallinga-Thie GM, Hautvast GAJ. Dependence of the effects of dietary cholesterol and experimental conditions on serum lipids in man. III. The effect on serum cholesterol of removal of eggs from the diet of free-living habitually egg-eating people. Am J Clin Nutr 1979;32:2193–7.
18. Schutz Y, Lechtig A, Bradfield RB. Energy expenditures and food intakes of lactating women in Guatemala. Am J Clin Nutr 1980;33:892–902.
19. Bowering J, Lowenberg RL, Morrison MA. Nutritional studies of pregnant women in East Harlem. Am J Clin Nutr 1980;33:1987–96.
20. Kay RM, Sabry ZI, Csima A. Multivariate analysis of diet and serum lipids in normal men. Am J Clin Nutr 1980;33:2566–72.
21. Kohrs MB, Nordstrom J, Plowman EL, et al. Association of participation in a nutritional program for the elderly with nutritional status. Am J Clin Nutr 1980;33:2643–56.
22. Wade J, Milner J, Krondi M. Evidence for a physiological regulation of food selection and nutrient intake in twins. Am J Clin Nutr 1981;34:143–7.
23. Marlett JA, Bokram RL. Relationship between calculated dietary and crude fiber intakes of 200 college students. Am J Clin Nutn 1981;34:335–42.
24. Anderson GH, Blendis LM. Plasma neutral amino acid ratios in normal man and in patients with hepatic encephalopathy: correlations with self-selected protein and energy consumption. Am J Clin Nutr 1981;34:377–85.
25. Naeye RL. Nutritional/nonnutritional interactions that effect the outcome of pregnancy. Am J Clin Nutr 1981;34:727–31.
26. Lee CJ, Lawler GS, Johnson GH. Effects of supplementation of the diets with calcium and calcium-rich foods on bone density of elderly females with osteoporosis. Am J Clin Nutr 1981;34:819–23.
27. Anderson BM, Gibson RS, Sabry JH. The iron and zinc status of long-term vegetarian women. Am J Clin Nutr 1981;34:1042–8.
28. King JC, Stein T, Doyle M. Effect of vegetarianism on the zinc status of pregnant women. Am J Clin Nutr 1981;34:1049–55.
29. Morgan KJ, Zabik ME, Leveille GA. The role of breakfast in nutrient intake of 5- to 12-year-old children. Am J Clin Nutr 1981;34:1418–27.
30. Clarke RP, Schlenker ED, Merrow SB. Nutrient intake, adiposity, plasma total cholesterol, and blood pressure of rural participants in the (Vermont) nutrition program for older Americans (Title III). Am J Clin Nutr 1981;34:1743–51.
31. Shorey RL, Bazan B, Lo GS, Steinke FH. Determinants of hypocholesterolemic response to soy and animal protein-based diets. Am J Clin Nutr 1981;34:1769–78.
32. Prentice AM, Whitehead RG, Roberts SB, Paul AA. Long-term energy balance in child-bearing Gambian women. Am J Clin Nutr 1981;34:2790–9.
33. Siegel S. Nonparametric statistics for the behavioral

sciences. Tokyo: McGraw-Hill Kogakusha, 1956:202–13.
34. Colton T. Statistics in medicine. Boston, MA: Little, Brown and Company, 1974:142–6.
35. Daniel WW. Biostatistics: A foundation for analysis in the health sciences. New York, NY: John Wiley and Sons, 1974:132.
36. Feinstein AL. The other side of statistical significance: alpha, beta, delta, and the calculation of sample size. Clin Pharmacol Ther 1975;18:491–505.
37. Fleiss JL. Statistical methods for rates and proportions. New York, NY: John Wiley and Sons, 1973:176–94.