European Journal of Clinical Nutrition (1994) **48**, 643–653
Received 23 December 1993; accepted 14 April 1994

The 24-hour recall for Senegalese weanlings: a validation exercise

M.-C. Dop[1], Ch. Milan[2], Cl. Milan[2] and A. M. N'Diaye[3]

[1]Laboratoire de Nutrition Tropicale, ORSTOM, 911 avenue Agropolis, Montpellier; [2]Registre Bourguignon des Cancers Digestifs, 7 boulevard J. d'Arc, Dijon, France and [3]Organisme de Recherche sur l'Alimentation et la Nutrition Africaines, 39 avenue Pasteur, Dakar, Senegal

Objective: In Africa, where growth retardation is highly prevalent, the use of expensive and time-consuming diet assessment techniques is a major obstacle to the collection of dietary data on large samples of children. The 24-h recall could be a valuable alternative. Its validity and reproducibility were assessed by comparison with the precise weighing technique.
Design & Subjects: Intakes of complementary foods of 45 Senegalese weanlings, aged 11–18 months, were estimated with both techniques on the same days.
Results: There was neither a level-dependent nor a systematic bias with the 24-h recall for energy and macronutrient intakes. Mean differences between techniques were <11% of mean intakes. Rank order correlations (r) ranged from 0.60 to 0.81 for energy and macronutrients. However, the analysis by food group showed that the 24-h recall was less precise than the reference, especially for foods from the household common pot (r ranged from 0.31 to 0.61 for rice, oil and fish). In particular, measuring rice intake as a number of 'handfuls' was not satisfactory. The technique needs further improvement for these foods.
Conclusion: Since it provides unbiased estimates of weanlings' intakes, the 24-h recall can be used for diet surveys of groups of children. Its lack of precision, which could cause attenuation in epidemiological studies, can be compensated by increasing the number of days of survey. This study suggests that a precision equivalent to that of a 1-day weighed survey could be obtained with two 24-h recalls, at a considerably lower cost.
Sponsorship: French Ministry of Research, grant no. 87G0517.
Descriptors: Africa, dietary methodology, 24-h recall, preschool children

Introduction

Two decades ago, Rutishauser (1973) noted that 'relatively few studies of food intake by preschool children have been made anywhere'. The lack of data was particularly critical for developing countries. In Africa, where growth retardation is still highly prevalent (Carlson & Wardlaw, 1990), little progress has been made since then in the development and validation of simple and inexpensive diet assessment methods for preschool children.

Children's intakes are small and therefore difficult to measure. An absolute error deemed reasonable in a survey of adults' intakes may represent an unacceptably high relative error in a survey of children. Consequently, most studies use detailed methods such as the precise weighing technique (PWT) or the collection of duplicate diets (Rutishauser, 1973). These methods are expensive and time-consuming. Therefore the number of subjects and/or the number of days of survey is generally small. Studies may thus fail to demonstrate a relationship between

Correspondence to: Dr Marie-Claude Dop.

the diet and growth retardation because of lack of power and/or attenuation due to intra-individual variation (Beaton *et al.*, 1979). Moreover, an overly cumbersome method might interfere with food intake; in such a case, measured intakes do not reflect habitual food consumption (Rutishauser, 1973; Isaksson, 1993). An additional difficulty, specific to many African countries, is the custom of hand-feeding children from the household common pot. Intake cannot be estimated by means of household measures. Separating the child's portion from the common pot in order to weigh it can alter the mother's way of feeding and consequently the child's intake.

In contrast to the detailed techniques, the 24-h recall (24-HR) is a quick and inexpensive method that could be used in the context of nutritional surveys of large groups of children. Since it estimates intakes over a short period of time, it could be used concurrently with growth monitoring to assess children's rapidly changing diets. When the goal is to measure individuals' usual intake, i.e. when it is necessary to assess intakes over more than one day, the 24-HR can be repeated with little inconvenience to the respondents.

The aim of this study was to assess the validity and reproducibility of the 24-HR for measuring Senegalese children's intakes of complementary foods during the weaning period. The study was part of a research programme on the determinants of growth retardation in Senegal.

Subjects and methods

Study design and location

The study took place in Pikine, the densely populated suburb of Dakar (capital of Senegal, West Africa). Reference techniques, against which the 24-HR was to be judged, were a qualitative observation of the children's food consumption and the PWT. The reference techniques and the 24-HR were used to estimate intakes on the same days (Fig. 1). For each child the study lasted 4 days: there was a day of qualitative observation (d0) and 2 days of precise weighing (d1 and d2). 24-HR interviews took place on d1, d2 and d3.

Subjects

The participants were children consuming foods other than breast-milk regularly (i.e. complementary foods). Families were contacted door-to-door in the vicinity of a local primary health care centre. The study was carried out in the children's homes. Parents were asked for their oral informed consent. Nutritional counselling was given after the survey.

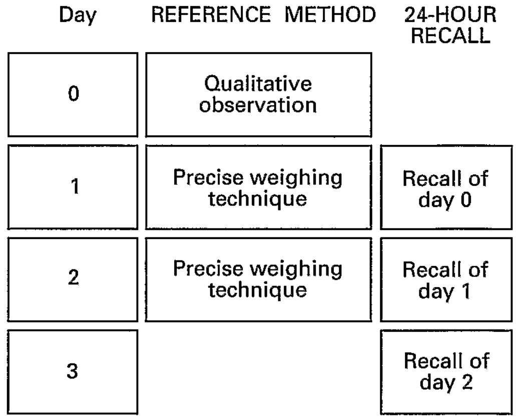

Fig. 1. Study design.

Forty-five children were recruited, 24 boys and 21 girls, aged 11–18 months. Thirty-eight were mixed-fed, i.e. received both breast-milk and complementary foods, and seven were fully weaned. Eight children presented minor illnesses such as a cold or diarrhoea during the study.

Diet assessment techniques

Qualitative observation. The child's qualitative food consumption was observed and recorded as unobtrusively as possible by a field-worker.

Precise weighing technique (PWT). This technique (also known as the 'observed weighed record') (Paul, 1988) was used as the reference for quantitative intakes. Raw ingredients, cooked dishes and the children's portions were weighed on household beam scales (precision 1 g, capacity 10 kg). It was necessary to adapt the technique to measure consumption from the household common pot; when the child was fed from the common pot, the field-worker fitted him with a disposable diaper, weighed him on an electronic portable scale (precision 1 g, capacity 30 kg, Ktron Inc.) before and after the meal, and recorded the number of mouthfuls he ate of each ingredient. Samples of cooked ingredients were analysed for humidity.

The reference techniques, i.e. the qualitative observation and the PWT, were implemented by three field-workers. Each field-worker stayed with a child from 7 a.m. until bedtime for three consecutive days (d0 to d2). At meal times the field-worker calibrated the mother's 'handfuls' of rice or millet from the household common pot.

Twenty four-hour recall (24-HR). The 24-HRs were conducted by an independent interviewer according to the standard methodology (Löken, 1988). For each meal, both the person(s) who prepared the food and who fed the child were interviewed. To estimate the intake of liquid or semi-liquid foods, the interviewer asked the respondent (often the mother) to reproduce the volume served as well as the leftover with water in the child's cup. The interviewer then quickly weighed the cup on a small portable electronic scale (precision 2 g, capacity 2 kg). The intake of rice and millet from the household common pot was estimated as a number of mothers' 'handfuls'. Weights of individual mothers' handfuls calibrated during the qualitative observation and PWT, and other calibrations performed before the study were pooled, and 'mean mothers' handfuls' were calculated for nine recipes. Thus two methods were used to obtain weights of foods from the number of handfuls recalled: one based on the weight of each mother's handfuls, the other based on the weight of 'mean handfuls' (the data subsequently presented in the Tables are based on 'mean handfuls'). The intake of fish was estimated as a number of 'fingers'. The average duration of an interview was 20 min. Mothers were not told in advance of the exact timing of the interviews so that they would not prepare themselves to recall their child's food intakes.

A previous survey performed in the same area, using the precise weighing technique, provided mean recipes, including the humidity of cooked ingredients (Dop *et al.*, 1994). They were used to translate intakes of cooked foods into amounts of raw ingredients. Before and during the food survey, information was collected from local markets, shops and street vendors on sizes, prices, recipes and the humidity of numerous foods and mixed dishes.

Great care was taken to avoid any exchange of information between the field-workers implementing the reference techniques and the interviewer conducting the recalls. Moreover, there were two independent supervisors, one for the reference techniques and one for the recall (the first author).

Data coding and analysis

The local food composition table (ORANA, 1982) and the table for use in Africa (FAO, 1970) were used to compute nutrient intakes with both the PWT and the 24-HR. Seven food groups representing 92% of energy intake were selected for presentation: wheat products, millet-sorghum, dairy products, added sugar, rice, oil and fish. Wheat products reflect the consumption of bread, pastries and cookies. Millet-sorghum and added sugar are most often ingredients of gruels. Dairy products are used in beverages and gruels. Rice, oil and fish are usually the main ingredients of the household common pot. Analyses of agreement between the techniques by food group were carried out with the log-linear model (three-way tables), paired *t*-tests and rank order correlations. For the computation of correlations, subjects who were non-consumers of a food group according to the reference technique, the PWT, were excluded because they were likely to be non-consumers with the recall, and zero values with both techniques would have improved agreement fictitiously.

The assessment of agreement in terms of energy and macronutrient intakes was done in three steps: the detection (i) of a level-dependent bias (the so-called 'flat slope syndrome') according to the method of Burema, Van Staveren & Van den Brandt (1988), (ii) of a systematic bias (paired *t*-tests), and (iii) a measure of the strength of the association (rank order correlations).

Estimates of ratios of intra- to inter-individual variance and attenuation with the PWT and the 24-HR were calculated according to the method of Liu *et al.* (1978):

$$\text{'error term'} = \sqrt{1/(1 + \sigma_a^2/n\sigma_r^2)} \qquad (1)$$

where σ_a^2 is the intra-individual variation, n the number of days of measurement and σ_r^2 the inter-individual variation. The ratio σ_a^2/σ_r^2 was estimated from the product-moment correlation r_{12} of intakes between d1 and d2 for each technique:

$$\sigma_a^2/\sigma_r^2 = (1 - r_{12})/r_{12} \qquad (2)$$

Estimates of 'error terms' were used to compare the attenuation due to the 24-HR to that of the PWT (an 'error term' close to 1 indicates a negligible attenuation of the true correlation between the diet and a disease variable of interest).

The 5% level was chosen for statistical significance. Results are expressed as means ± SE. Correlation coefficients are given with their 95% confidence intervals (CI).

Results

Analysis by food group

Qualitative agreement between the techniques, based on whether or not a food group was mentioned during a day with each technique, is shown in Table 1. There are four categories of agreement: column (1) represents children who consumed a food according to both the reference technique (qualitative observation or PWT) and the 24-HR, (2) children who consumed a food according to the reference but the food was omitted with the 24-HR, (3) children who did not consume a food according to the reference but the food was added with the 24-HR, (4) children who did not consume a food according to both the reference and the 24-HR. Day 0 represents the comparison of the 24-HR with a day of qualitative observation, while d1 and d2 represent the comparison with the PWT.

Most errors were omissions with the 24-HR; the number of children for whom a food group was omitted was less than five for all food groups except fish. Omissions as a proportion of consumers exceeded 10% only for two food groups, millet-sorghum and fish. The highest number of omissions was found with fish on d2 (eight out of 34 consumers, 24%). The number of additions did not exceed two for all foods except fish, for which the highest number was observed on d2 (seven additions out of 20 non-consumers, 35%).

Table 1. Qualitative agreement between techniques for intakes of selected foods

			Number of children for whom a food group was:			
Food group	*Day*	*Comparison: 24-HR vs*	*Mentioned with both techniques (1)*	*Mentioned with the reference omitted with the 24-HR (2)*	*Unmentioned with the reference added with the 24-HR (3)*	*Unmentioned with both techniques (4)*
Wheat products	0	Observation	39	1	2	3
	1	PWT	40	1	0	4
	2	PWT	35	3	0	7
Millet-sorghum	0	Observation	21	4	0	20
	1	PWT	15	2	1	27
	2	PWT	17	2	1	25
Dairy products	0	Observation	33	2	0	10
	1	PWT	31	3	0	11
	2	PWT	32	1	0	12
Added sugar	0	Observation	44	1	0	0
	1	PWT	43	1	0	1
	2	PWT	41	0	2	2
Rice	0	Observation	44	0	0	1
	1	PWT	42	1	1	1
	2	PWT	44	0	1	0
Oil	0	Observation	45	0	0	0
	1	PWT	40	4	0	1
	2	PWT	41	1	0	3
Fish	0	Observation	26	8	1	10
	1	PWT	26	3	4	12
	2	PWT	21	4	7	13

Columns (1) and (2) are consumers, (3) and (4) are non-consumers according to the reference techniques.
24-HR = 24-h recall; PWT = precise weighing technique.

Table 2. Mean differences (±SE) between the 24-h recall (24-HR) and the precise weighing technique (PWT) for intakes of selected foods (24-HR – PWT) (n = 45)

Food group (g/day)	*Day 1*				*Day 2*			
	PWT	*24-HR*	*Difference*	*P*[a]	*PWT*	*24-HR*	*Difference*	*P*[a]
Wheat products	22.5 ± 2.4	18.3 ± 2.1	–4.2 ± 1.5	0.01	19.8 ± 2.4	19.1 ± 2.4	–0.7 ± 1.8	0.69
Millet-sorghum	13.4 ± 3.4	13.2 ± 3.4	–0.2 ± 1.7	0.92	15.0 ± 3.6	12.6 ± 3.0	–2.4 ± 1.5	0.12
Dairy products	8.6 ± 2.1	10.1 ± 2.6	+1.5 ± 0.9	0.12	6.6 ± 1.8	7.7 ± 1.8	+1.1 ± 0.8	0.19
Added sugar	18.5 ± 2.5	20.0 ± 2.1	+1.5 ± 1.5	0.34	17.2 ± 2.5	19.3 ± 2.2	+2.0 ± 1.2	0.11
Rice	29.9 ± 2.9	33.1 ± 3.0	+3.2 ± 3.3	0.34	29.7 ± 3.5	32.1 ± 2.9	+2.4 ± 3.3	0.47
Oil	5.2 ± 0.7	5.5 ± 0.7	+0.3 ± 0.7	0.69	4.3 ± 0.6	4.8 ± 0.6	+0.5 ± 0.6	0.38
Fish	10.9 ± 2.5	5.4 ± 1.1	–5.4 ± 2.3	0.02	6.5 ± 1.6	4.2 ± 0.8	–2.3 ± 1.2	0.06

[a] Paired *t*-tests.

The log-linear model was applied to three categories of agreement (columns 1, 2 and 3 of Table 1), the 3 days of survey and five food groups (dairy products and oil were excluded because of frequencies equal to zero in column 3). Agreement differed significantly between food groups ($P < 0.0001$) but not between days of survey ($P = 0.15$).

Quantitative agreement between the PWT and the 24-HR is presented in Table 2. Only wheat products and fish on d1 were significantly underestimated by the 24-HR. Fish intake was not significantly underestimated on d2 but the *P* value is near significance (0.06). The mean difference between techniques relative to the mean intake according to the PWT, was smaller for wheat products (–19%) than for fish (–50% on d1 and –35% on d2).

Rank order correlations are given in Table 3; they were high over both days of survey with millet-sorghum, dairy products and added sugar. They were intermediate, also over both days, with wheat products and oil. With rice and fish, they increased, from a low value on d1 to an intermediate value on d2, although not significantly.

Table 3. Rank order correlations between the 24-h recall (24-HR) and the precise weighing technique (PWT) for intakes of selected foods

Food group	*Day 1*		*Day 2*		*Mean of days 1 and 2*	
	n[a]	*Spearman's r (95% CI)*	*n*[a]	*Spearman's r (95% CI)*	*n*[a]	*Spearman's r (95% CI)*
Wheat products	41	0.63 (0.40–0.78)	38	0.59 (0.33–0.77)	41	0.76 (0.59–0.87)
Millet-sorghum	17	0.81 (0.53–0.93)	19	0.87 (0.68–0.95)	23	0.91 (0.79–0.96)
Dairy products	34	0.83 (0.68–0.91)	33	0.82 (0.66–0.91)	38	0.89 (0.80–0.94)
Added sugar	44	0.82 (0.70–0.90)	41	0.78 (0.62–0.88)	45	0.82 (0.69–0.90)
Rice	43	0.38 (0.09–0.61)	44	0.61 (0.38–0.77)	45	0.52 (0.26–0.70)
Oil	44	0.58 (0.34–0.75)	42	0.56 (0.31–0.74)	45	0.59 (0.35–0.75)
Fish	29	0.31 (–0.06–0.61)	25	0.56 (0.21–0.78)	34	0.43 (0.11–0.67)

[a] Non-consumers according to the PWT are excluded.

Children frequently consumed rice, but seldom millet, from the household common pot. Thus, the food most often measured as 'handfuls' was rice. The intake of oil was derived from that of rice according to mean recipes. Therefore estimations based on 'handfuls' were used almost exclusively for rice and oil. When each mother's 'handfuls' were used to compute intakes of these two foods, instead of 'mean handfuls' (given in Table 2), mean rice intakes of d1 and d2 were respectively 31.8 ± 2.9 and 31.3 ± 2.9 g, and mean oil intakes 5.2 ± 0.6 and 4.6 ± 0.5 g. Differences between the PWT and the 24-HR were only negligibly smaller, *P* values of paired *t*-tests were very similar and rank order correlations were unchanged.

Energy and macronutrient intakes

The mean energy and macronutrient intakes measured with the PWT and the 24-HR, mean differences between the techniques and regression coefficients for the detection of a level-dependent bias are given in Table 4. Regression coefficients of the difference between observations of each technique on the sum of the observations did not differ significantly from zero. In the absence of a level-dependent bias, mean differences were used to detect a systematic bias. The largest mean difference relative to the mean intake given by the PWT, was observed with protein intake on d1 (–10.6%). When protein intakes of d1 and d2 were averaged, the mean difference between techniques was smaller (–7.2%, $P = 0.16$). For the other macronutrients the mean difference was smaller than ± 7% on both d1 and d2. Paired *t*-tests failed to detect a systematic bias. When d1 and d2 were compared, mean macronutrient intakes appeared lower on d2 with both techniques, but the decrease was significant only for protein and fat intake with the PWT ($P = 0.02$). For energy with the PWT, the difference was near significance ($P = 0.08$).

Rank order correlations for nutrient intakes are shown in Table 5. Although they appeared higher on d2, they did not differ significantly because of the large confidence intervals. Correlations computed from mean intakes of d1 and d2 were all 0.70 or more.

When individual mothers' 'handfuls' were used instead of 'mean handfuls', protein and fat intakes given by the 24-HR were unchanged while mean energy intake was higher on d1, and

Table 4. Mean differences (±SE) between the 24-h recall (24-HR) and the precise weighing technique (PWT) for nutrient intakes (24-HR – PWT) ($n = 45$)

Nutrients	Day 1					Day 2				
	PWT	*24-HR*	*Difference*	P^a	β^b	*PWT*	*24-HR*	*Difference*	P^a	β^b
Energy (kJ/day)	1773 ± 119	1769 ± 123	–4 ± 98	0.97	+0.016	1634 ± 117	1690 ± 108	+56 ± 90	0.53	–0.048
Protein (g/day)	10.4 ± 0.9	9.3 ± 0.8	–1.1 ± 0.7	0.13	–0.012	8.9 ± 0.9	8.5 ± 0.8	–0.4 ± 0.5	0.45	–0.068
Fat (g/day)	11.5 ± 1.0	11.2 ± 1.0	–0.3 ± 0.9	0.73	–0.014	9.7 ± 1.0	10.3 ± 1.0	+0.6 ± 0.7	0.41	+0.011
Carbohydrate (g/day)	70.4 ± 5.0	72.0 ± 5.0	+1.6 ± 3.7	0.68	–0.006	67.6 ± 5.0	70.0 ± 4.3	+2.4 ± 3.6	0.51	–0.092

[a] Paired *t*-tests.
[b] Regression coefficient of d = 24-HR – PWT on s = PWT + 24-HR (Burema *et al.*, 1988).

Table 5. Rank order correlations between the 24-h recall (24-HR) and the precise weighing techniques (PWT) for nutrient intakes (n = 45)

	Day 1	*Day 2*	*Mean of days 1 and 2*
Nutrients	*Spearman's r (95% Cl)*	*Spearman's r (95% Cl)*	*Spearman's r (95% Cl)*
Energy	0.63 (0.42–0.78)	0.78 (0.64–0.88)	0.75 (0.59–0.86)
Protein	0.66 (0.45–0.80)	0.79 (0.65–0.88)	0.75 (0.58–0.85)
Fat	0.60 (0.37–0.76)	0.73 (0.55–0.84)	0.70 (0.51–0.82)
Carbohydrate	0.65 (0.45–0.79)	0.81 (0.68–0.89)	0.80 (0.67–0.89)

lower on d2, by only 20 kJ. Mean carbohydrate intakes were lower on both days by <1 g. P values for paired t-tests were very similar. Rank order correlations between the techniques were unchanged or only negligibly higher (by at most 0.03). Correlations for mean intakes of d1 and d2 were changed by <±0.04.

Estimated 'error terms' for energy and macronutrient intakes of a 1-day survey calculated from formulae (1) and (2) (with n = 1), ranged from 0.86 (fat) to 0.90 (carbohydrate) with the PWT, and from 0.75 (fat) to 0.87 (protein) with the 24-HR. Again, when each mother's 'handfuls' were used, instead of 'mean handfuls', results were similar.

Discussion

Two concepts summarize the quality of a diet assessment technique: validity, which refers to systematic bias, and reproducibility, which refers to random measurement error. This approach guided our analysis. Moreover, there were two levels of analysis, foods and nutrients. The detailed analysis by food group enabled us to understand how the technique performed for each type of food, and in particular which foods were difficult to estimate by recall. The overall quality of the 24-HR was judged in terms of nutrient intakes.

Subjects

Because of problems of access by car, the subjects were not selected in the whole population of Pikine and the area of the study was restricted to the vicinity of a primary health care centre. A previous nutritional survey has shown that ethnic composition, socioeconomic status and nutritional status in the area chosen for the study do not differ from other parts of Pikine (Maire *et al.*, 1989). We chose not to exclude subjects presenting minor illnesses (as long as they consumed some complementary food during the survey) because a diet assessment technique to be used for estimating intakes of a population of children must be valid for ill as well as healthy children.

Statistical methods

Statistical methods were chosen for their computational simplicity and unambiguous meaning. Pearson correlation coefficients, intraclass correlations and regression coefficients of one technique on the other have been criticized by several authors (Altman & Bland, 1983; Burema *et al.*, 1988; Borrelli, Cole & Contaldo, 1989; Chinn, 1990). One of their limitations is that they are influenced by the range of values, i.e. the amount of between-subject variation found in the sample. A large between-subject variation will tend to produce a high correlation whatever the level of true agreement between the techniques. Rank order correlations, on the contrary, do not suffer from this limitation because they are based on ranks of intakes. Moreover they do not depend on the choice of a number of categories, as the Kappa statistic does, while conveying essentially the same information (Maclure & Willett, 1987). Regression coefficients of the diet assessment technique to be validated on the reference technique are affected by the phenomenon of 'regression to the mean' and therefore cannot be used to detect a level-dependent bias. Burema *et al.* (1988) have proposed a simple method for the detection of this bias. When the difference between techniques is larger (negative or positive) at the higher and lower ends of the distribution of intakes, the regression coefficient of the difference between the techniques on the sum of the techniques differs from zero. This type of bias is usually in the direction of an overestimation at the lower end of the distribution and an underestimation at the higher end, producing a negative regression coefficient. The absence of a level-dependent bias is a prerequisite for the detection of a systematic bias, the latter being meaningless if the difference between techniques is not constant.

Problems of the reference technique

The choice of a reference technique for the validation of a diet assessment method is a difficult problem. A non-dietary reference, for example a biological marker, is preferable because it is independent of diet assessment (Willett, 1990). In the present study, since only the intake of complementary foods was of interest (and not breast-milk) only another diet assessment technique could serve as the reference.

The PWT was chosen rather than the collection of a duplicate diet because we feared the latter might help mothers to memorize their child's intake and thus invalidate the comparison. For the purpose of validation, the precision of the reference is an important issue, because it affects the power of the comparison through the variance of differences. In order to ensure the highest precision possible for the reference, each field-worker carrying out the PWT surveyed only one child at a time. Field-workers were instructed to stay with the child all day so that all foods would be weighed, unlike the method used in Keneba (The Gambia) where field-workers were present only at meal times and snack consumption was estimated by recall (quoted by Paul, 1988).

In Dakar as in Keneba, the PWT is confronted with the problem of hand-feeding from the household common pot. Unlike Le François & Chevassus-Agnès (1983), we chose not to ask mothers to separate the child's portion from the common pot in order to weigh it, so as not to help them remember the child's intake. The technique of weighing the child before and after the meal and calculating intake from the number of mouthfuls was chosen despite its lower precision because it was the least intrusive, and thus less likely to cause a memory bias.

The analysis of macronutrient intakes by day of survey showed a tendency towards reduced intakes on d2. This could be due to an 'instrument effect' of the PWT. Such an effect was observed in a previous study carried out in the same community, where a 7-day weighed survey caused a significant decrease in the children's food intake and weight-gain (Dop *et al.*, 1994); because of the 'instrument effect', intakes measured during the study did not reflect the subjects' usual food consumption. In the present study, if the diet on which the validation is based were different from the usual diet, the validity of the 24-HR might not extend to the children's usual diet. However, qualitative intakes did not vary with the day of survey, i.e. whether intakes were observed (d0) or weighed (d1 or d2); thus intakes were not affected qualitatively by the 'instrument effect' of the PWT. Moreover, the mean reduction of intakes between d1 and d2 was 15% at most. The 'instrument effect' caused only a moderate quantitative reduction in the children's intakes. Hence a similar validity can be expected of the 24-HR for future diet surveys of Senegalese weanlings.

Validity of the comparison

When the reference and the technique to be validated are used on the same days of intake, it is particularly important to make sure that sources of error are not correlated, i.e. that the implementation of the reference technique does not improve the results of the technique to be validated (Rutishauser, 1973; Willett, 1990). Actually many studies fail to address this problem. We attempted to examine it by including in our study a day of qualitative observation, with no weighing; this enabled us to compare the qualitative agreement whether the 24-HR followed a day of observation or a day of PWT. Qualitative agreement did not improve when the 24-HR followed the PWT, indicating that the techniques were independent. Nevertheless this analysis was not very sensitive, as it was based on qualitative data, and only an important departure from independence could have been detected.

Problems with the structure of data
Some aspects of the structure of data coding limited the depth of the analysis. Intake was coded by meal and not by type of dish. With the PWT, the meal was defined according to the time of consumption; for instance breakfast was defined as all food eaten between waking up and 10 a.m. This definition often did not correspond to the meal pattern of mothers, and besides they were not able to recall the time of meals. Therefore a comparison by meal of the PWT and the 24-HR was not relevant.

With the 24-HR, the method of quantification used by the interviewer depended on the form under which the food was acquired and used, for example whether it could be priced or whether a household utensil was used, etc. In other words, the method of quantification was dish-specific. When neither pricing nor household utensils could be used, as in the case of hand-fed foods, quantification was obviously more difficult. An analysis by dish would have enabled us to judge the adequacy of each of these methods. The analysis by food group was used as a substitute for the type of dish because most foods were used in only one type of dish. Nevertheless there were exceptions. For example, millet, an ingredient of gruels, can also be used in the household common pot.

Qualitative agreement
The qualitative analysis was not intended to assess, *per se*, the level of agreement between the techniques. Moreover for some food groups, it is possible that it overestimated true agreement. For foods that are commonly consumed more than once a day, if an omission occurred with the 24-HR, it went undetected because agreement was judged on a dichotomous basis (mentioned or unmentioned). An analysis of frequency of consumption during the day should have been done instead. However, in this study, frequency was not relevant because mothers, during interviews, tended to group intakes of several snacks of the same dish. For example, when a mother proposed breakfast leftovers of millet-gruel to her child several times during the rest of the day, she found it easier to recall the total amount consumed during the day than to describe each separate snack.

The number of qualitative errors differed widely by food group. It was smallest for rice and added sugar and highest for fish. Most errors were omissions, except for fish which showed a high number of both omissions and additions. Some of the omissions of fish may not have been due to the 24-HR but to the PWT. Mothers usually broke up pieces of fish over the rice and it was difficult for the PWT fieldworkers to distinguish these two foods. Consequently, the intake of fish may have been overestimated with the PWT, resulting in a fictitious underestimation with the 24-HR. Apart from fish, the highest number of errors was found with millet-sorghum, mostly omissions. They had little impact on quantitative agreement, i.e. they did not result in a significant underestimation. It seems that mothers neglected to mention millet-sorghum when the quantity consumed by the child was small.

Quantitative agreement
In terms of the absolute intake of foods, there was no systematic bias of the 24-HR consistent over both days of survey (d1 and d2) for most food groups. Fish was the exception, with an important underestimation on both d1 and d2. For the main ingredients of the household common pot, rice, oil and fish, SEs of differences between the techniques were large in comparison with SEs of each technique. This shows that mothers had difficulty in quantifying their child's intake of these foods and suggests that their estimation was affected by a large random error.

The problem of quantification is well reflected, in terms of relative intake, by rank order correlations: the 24-HR ranked subjects correctly for foods served or fed in household measures or as standard portions (wheat products for example); beverages and gruel were well quantified by asking mothers to reproduce the child's portion with water (dairy products, millet-sorghum). The major problem remains the estimation of foods from the household common pot for which neither household utensils nor standard portions can be relied upon. Rice and oil, which contribute 39% of the children's energy intake from complementary foods, come almost exclusively from the common pot. Measuring these foods as a number of 'handfuls' was not satisfactory. Using each mother's 'handfuls' instead of 'mean handfuls' only negligibly improved the results. Thus error in intakes recalled as 'handfuls' was not due to differences between mothers, it was a component of random measurement error. As shown by Liu *et al.* (1978), measuring intakes over several days reduces the amount of random error. Thus the precision of this estimation could be improved by increasing the number of days of survey.

The calibration of each of the survey mothers' 'handfuls', which can be costly and time-consuming, is not necessary. Calibration of 'handfuls' can be done before the start of a survey and a mean value used for all the children. Nevertheless, in future studies, alternative methods for measuring the intake of rice should be explored. The use of food models or actual samples of cooked dishes, which the mothers could measure out, should be experimented.

Energy and macronutrient intakes

In this study, a high level of agreement between the 24-HR and the PWT was expected because both techniques were used on the same days. The underestimation of group energy and macronutrient intakes, observed with the 24-HR in a majority of the studies reviewed by Bingham (1987) was not found here. There was no level-dependent bias, i.e. there was neither an under- nor an overestimation of intakes at the higher and lower ends of the distribution. Overall, there was no systematic bias for individual energy and macronutrient intakes. However, as with intakes of foods, the SEs of differences between techniques were large, of the same magnitude as SEs of each technique; nutrient intakes estimated with the 24-HR were affected by a large random error. For the purpose of epidemiological studies, the presence of random measurement error is not a major obstacle, as long as the method is not expensive, because it can be compensated by increasing the number of days of survey (Beaton *et al.*, 1979). The gain in precision per additional day can be calculated from formula (1) (Liu *et al.*, 1978). If a 2-day survey were carried out, 'error terms' with the 24-HR calculated with $n = 2$ in formula (1) would range from 0.86 (fat) to 0.93 (protein), reaching the level observed with a 1-day survey with the PWT.

There are few published validation studies of the 24-HR for preschool children from Africa. Rutishauser (1973) compared the mean of seven 24-HRs per child to the analysis of duplicate diets. She found no systematic bias with the 24-HR, but SEs of the differences between methods were larger than in our study. The methods were not used to assess intakes over exactly the same time period and differences between methods reflect both measurement error and intra-individual variation. The study design of Ferguson *et al.* (1989) from Malawi is more comparable to ours. The authors also found no systematic bias of the 24-HR in comparison with the weighing technique but individual agreement was low. This could be a consequence of the older age of the subjects (4–6 years): the children probably ate by themselves, and were not constantly under their mother's supervision, as they were in our study.

Conclusion

In Dakar, the 24-HR is a valuable method for assessing weanlings' group macronutrient intakes. It is not affected by level-dependent or systematic bias but it is less precise than the weighing technique. This is not a major limitation for its use in epidemiological studies because the technique is relatively inexpensive and its precision can be improved by increasing the number of days of survey. This study suggests that two 24-HRs should be performed in order to obtain a precision approximately equivalent to a 1-day survey with the precise weighing technique. While one field-worker using the 24-HR can easily survey eight to ten children in one working day, provided the distance between homes is not too large, a field-worker using the precise weighing technique can only survey one child. Thus, even if the number of days of survey were doubled, the advantage of the 24-HR in terms of time and cost would still be considerable. With the 24-HR, food surveys of large groups of preschool children could be carried out at an acceptable cost. The technique could prove to be a valuable tool for epidemiological studies of the dietary determinants of growth retardation in Africa.

Designing and analysing a validation study is a complex task. At various stages of the study, choices are to be made which will inevitably bear on the interpretation of the results. Issues such as the precision of the reference technique, the independence of techniques, the choice of a coding method to permit meaningful comparisons, the relevance to the usual diet, i.e. whether conclusions will apply to the population at large, deserve special attention. Our study is an illustration of how these issues can be addressed. In this respect, this 'validation exercise' can be of interest to researchers working in other contexts.

Acknowledgements—We thank the children's mothers for their kind cooperation, Marie-Anne Tressol for her efficient supervision, Ngoné Senne for the laboratory work and Marguerite Guiguet for her valuable methodological advice.

References

Altman DG & Bland JM (1983): Measurement in medicine: the analysis of method comparison studies. *The Statistician* **32**, 307–317.

Beaton GH, Milner J, Corey P, McGuire V, Cousins M, Stewart E, de Ramos M, Hewitt D, Grambsch PV, Kassim N & Little JA (1979): Sources of variance in 24-hour dietary recall data: implications for nutrition study design and interpretation. *Am. J. Clin. Nutr.* **32**, 2546–2559.

Bingham SA (1987): The dietary assessment of individuals; methods, accuracy, new techniques and recommendations. *Nutr. Abstr. Rev. (A)* **57**, 705–742.

Borrelli R, Cole TJ & Contaldo F (1989): Some statistical considerations on dietary assessment methods. *Eur. J. Clin. Nutr.* **43**, 453–463.

Burema J, Van Staveren WA & Van den Brandt PA (1988): Validity and reproducibility. In *Manual on methodology for food consumption studies*, eds ME Cameron & WA Van Staveren, pp. 171–182. Oxford: Oxford University Press.

Carlson BA & Wardlaw T (1990): A global, regional and country assessment of child malnutrition. Staff working papers, no. 7. New York: UNICEF.

Chinn S (1990): The assessment of methods of measurement. *Stat. Med.* **9**, 351–362.

Dop MC, Milan Ch, Milan Cl & N'Diaye AM (1994): The multiple-day weighed record for Senegalese children during the weaning period: a case of the 'instrument effect'. *Am. J. Clin. Nutr.* **59**, 266S–268S.

FAO (1970): *Table de composition des aliments à l'usage de l'Afrique*. Rome: Food and Agricultural Organization of the United Nations.

Ferguson EL, Gibson RS, Ounpuu S & Sabry JH (1989): The validity of the 24-hour recall for estimating the energy and selected nutrient intakes of a group of rural Malawian preschool children. *Ecol. Fd Nutr.* **23**, 273–285.

Isaksson B (1993): A critical evaluation of the duplicate-portion technique in dietary surveys. *Eur. J. Clin. Nutr.* **47**, 457–460.

Le François P & Chevassus-Agnès S (1983): Alimentation complémentaire de 38 nourrissons sénégalais. *Ann. Nutr. Metab.* **27**, 199–204.

Liu K, Stamler J, Dyer A, McKeever J & McKeever P (1978): Statistical methods to assess and minimize the role of intra-individual variability in obscuring the relationship between dietary lipids and serum cholesterol. *J. Chron. Dis.* **31**, 399–418.

Löken EB (1988): Twenty-four hour recall. In *Manual on methodology for food consumption studies*, eds ME Cameron & WA Van Staveren, pp. 53–106. Oxford: Oxford University Press.

Maclure M & Willett WC (1987): Misinterpretation and misuse of the Kappa statistic. Am. J. Epidemiol. **126**, 161–169.

Maire B, Chevassus-Agnes S, Griere B, N'Diaye T, Benart C & Salem G (1989): Etat nutritionnel des enfants d'âge préscolaire à Pikine. In *Urbanisation et santé dans le tiers monde*, eds G Salem & E Jeannee, pp. 105–113. Paris: coll. Colloques et séminaires, éditions de l'ORSTOM.

ORANA (1982): *Table de composition des aliments pour l'Afrique de l'Ouest.* Dakar: Organisme de Recherches sur la Nutrition et l'Alimentation Africaines.

Paul AA (1988): Observed weighed records. In *Manual on methodology for food consumption studies*, eds ME Cameron & WA Van Staveren, pp. 75–79. Oxford: Oxford University Press.

Rutishauser IHE (1973): Food intake studies in pre-school children in developing countries: problems of measurement and evaluation. *Nutrition* **27**, 253–261.

Willett W (1990): Reproducibility and validity of food-frequency questionnaires. In *Nutritional epidemiology*, pp. 92–126. New York: Oxford University Press.